Interlanguage of Japanese learners of English:
Judgements on the translatability of two polysemous
Japanese lexemes

Michiko Nakano

Ph. D.
Department of Applied Linguistics
University of Edinburgh
1987
Abstract

This research is concerned with the question of how we can explain the judgemental process of translatability, and the issue of whether or not research into the translatability judgements can illuminate learning processes, as Kellerman's experiments have suggested (1977, 1978 and 1979). The experiments deal mainly with two polysemous Japanese verbs 'kowasu' (break) and 'yaburu' (tear), involving five groups of Japanese learners of English. The experimental methods are intended to be "learner-oriented" and "process-oriented", in accordance with a research aim of inter-language. The experiments reported below try to demonstrate that the judgemental process is best seen as an interaction between data-processing and the analogical process. Both data-processing and the analogical process are known to be an important part of language learning or acquisition. For this reason, the present research provides an explanation not only for translatability judgements but also learning (or acquisitional) processes. Furthermore, the present interactive view of the judgemental process accords with Corder's view of language acquisition which consists of the three stages of data-processing, hypothesis-formation and hypothesis-testing (Corder, 1979). The present study demonstrates that the judgemental process of translatability involves all these stages.

The present research also provides a definition of lexical structure based upon Rosch's notional characterization (e.g., Rosch, 1975b and Mervis and Pani, 1980). The lexical structure is defined in terms of Boolean algebras. This tentative definition is useful in distinguishing 'competence' from 'performance' and ties up with the five types of data-processing discussed in chap. 9.

A learner has some 'sure' items whose translatability he is confident of. The translatability judgement is based upon an analogical process where a learner's 'feel' of similarity between his 'sure' item and the items he is judging is crucial. The function of 'sure' items is that of a prototype in the case of an L1 child. At the initial exposure to a novel NL lexeme, a learner induces not only a prototype, but also the limit of application of a novel lexeme. On later occasions he uses the already acquired projective limit to the other NL words. The present study adopts the Thurstone-Torgerson method to estimate inductive and projective limits. The experiments in chap. 6 & 7 (cf. appendix 5) confirm that the present application of the Thurstone-Torgerson method is adequate.

The judgemental process also involves iterative reorganization of NL interlexical structure which is initiated by the three principles of similarity (semantic closeness, maximal similarity and identity). This iterative reorganization is a reflection of categorical processing which is represented by Boolean operations. The present research offers evidence for the importance of Boolean operations in the judgemental process of translatability.
I should like to acknowledge a deep sense of gratitude to Dr. Alan Davies and Dr. S. Pit Corder, my supervisors at the department of Applied Linguistics. I am greatly indebted for their valuable supervision. With their guidance, I have managed to complete this thesis.

I should like to express my particular gratitude to Professor R. Asher for his many instances of sympathetic understanding and advice, and to Mr. Alan Kemp for his encouragement and concern.

I should like to thank Mr. Tony Howatt, Mr. Keith Mitchell and Professor Jim Hurford for their support and patience.

I should also like to express my gratitude to Mr. Norman Dryden, Mrs. Cynthia Dryden, Mrs. Irene Mcleod, Mrs. Syllvia Motherwell for their help and support.

I would like to express my gratitude to Mr. McMullen for his support, and to Mrs. Parker for her encouragement.

I should also like to express my thanks to Professor Togari, Professor Kanatani, Ms. Kaeko Mochizuki, Mr. Mukai, Mr. Watanabe, and Mr. Ito for arranging experiments at Ehime University and for giving me their expert advice and hospitality. I should also like to thank Ms. Sinohara for arranging experiments at Minami High School. I would like to express my gratitude to Mr. Gichi Watanabe for arranging experiments at Shonan Middle school.

I should also like to express my thanks to Dr. Tamio Arai (Tokyo University), Ms. Sagara and Mr. Kakiuchi (Aichi University) for their mathematical advice.

I am also grateful to Dr. Ikuo Koike (Keio University and the vice chairman of JACET), Dr. Seiji Shibata (Naruto Graduate School of Education) and Dr. Ineko Kondo (Tsuda College) for their support and advice.
I would like to thank Ms. Sherri Scott, Dr. I. Maxwell Hammett and Mr. Ivan Cosby for their enthusiasm and encouragement.

I am also grateful to Yoji Tamakoshi, Mr. Shintani, Mrs. Takekawa and my students for their typing.

This thesis is composed entirely on my own, but because many friends and colleagues of mine including my supervisors have charitable minds, I have managed to finalize it this time.

I am also grateful to my colleagues at Aichi University and Edinburgh University for their encouragement and friendship.
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Contents</td>
<td>vi</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>x</td>
</tr>
<tr>
<td>Chapter 1: introduction</td>
<td>1</td>
</tr>
<tr>
<td>§1.0</td>
<td>1</td>
</tr>
<tr>
<td>§1.1 basic terms in IL studies</td>
<td>3</td>
</tr>
<tr>
<td>§1.2 thesis outline</td>
<td>11</td>
</tr>
<tr>
<td>Chapter 2: selective review of literature</td>
<td>22</td>
</tr>
<tr>
<td>§2.0 introduction</td>
<td>22</td>
</tr>
<tr>
<td>§2.1 the previous methodology: performance analysis</td>
<td>23</td>
</tr>
<tr>
<td>§2.2 learner-language and process-oriented research</td>
<td>30</td>
</tr>
<tr>
<td>Chapter 3: some preliminaries adopted in the present study</td>
<td>43</td>
</tr>
<tr>
<td>§3.0 introduction</td>
<td>43</td>
</tr>
<tr>
<td>§3.1 three hypotheses in L\textsubscript{1} lexical acquisition</td>
<td>47</td>
</tr>
<tr>
<td>§3.2 previous experimental works on PH</td>
<td>57</td>
</tr>
<tr>
<td>§3.3 previous two studies on analogical reasoning</td>
<td>65</td>
</tr>
<tr>
<td>§3.4 an operational definition of word-meaning</td>
<td>71</td>
</tr>
<tr>
<td>§4.0 purpose</td>
<td>88</td>
</tr>
<tr>
<td>§4.1 summary and comment: Kellerman's study</td>
<td>94</td>
</tr>
<tr>
<td>§4.2 operational definitions of three types of processing</td>
<td>97</td>
</tr>
<tr>
<td>§4.3 a pilot experiment</td>
<td>102</td>
</tr>
</tbody>
</table>
§4.3.1 materials used 102
§4.3.2 the method of elicitation of input data for MDS and for the Thurstone-Torgerson method 106
§4.3.3 the elicitation method of translatability judgements 107
§4.4 results and discussion 110
§4.4.1 analysis I 110
§4.4.2 analysis II 128
§4.4.3 analysis III 148
§4.5 summary 153
§4.6 plan of research 155

Chapter 5: pilot study (2);
materials, the methods of rating and MDS elicitation 158
§5.0 purpose 158
§5.1 the method of choosing object nouns 158
§5.1.1 experiment 161
§5.1.2 results 162
§5.2 the number of steps in the rating scale 166
§5.2.1 experiment 171
§5.2.2 results 172
§5.3 the method of elicitation for MDS 175
§5.3.1 a brief overview of MDS 175
§5.3.2 a brief overview of elicitation methods of the MDS input 179
§5.3.3 categorical rating (sorting) 184
§5.3.4 the treatment of individual scores 185

Chapter 6: pilot study (3); free recall experiments
§6.0 purpose 190
§6.1 overview 191
§6.2 experiment (1) 198
§6.2.1 procedure 199
§6.2.2 results and discussion 200
§6.2.3 summary 202
§6.3 experiment (2)
§6.3.1 procedure 205
§6.3.1.1 the method of presentation 206
§6.3.1.2 administration 208
§6.3.2 the method of counting clusterings 209
§6.3.3 results and discussion 212
§6.3.4 summary 220

Chapter 7: pilot study (4); experiments concerning the application of the Thurstone-Torgerson method
§7.0 purpose 222
§7.1 overview 225
§7.2 experiments
§7.2.1 hypothesis 1 236
§7.2.2 administration: experiment 1 239
§7.2.3 result: experiment 1 240
§7.2.4 hypothesis 2 241
§7.2.5 administration: experiment 2 245
§7.2.6 results and discussion: experiment 2 246
§7.3 conclusion 248
§7.4 an application of SIM to the analysis of translatability judgement data 249

Chapter 8: pilot study (5); experimenting the hypothesis of compound vs coordinate bilinguals
§8.0 purpose 253
§8.1 overview 254
§8.1.1 a specification of previous experimental areas 257
§8.1.2 some experimental difficulties involving L2 subjects 267
§8.1.3 selective review 272
§8.2 experiments 276
§8.2.1 administration 276
§8.2.2 results and discussion 277
Chapter 9: pilot study (6); experiments to determine
a form of processing

§9.0 purpose 283
§9.1 another use of MDS 285
§9.2 the relationship between processings
and metrics 291
§9.3 experiments 314
  §9.3.1 method 314
  §9.3.2 results and discussion 317
  §9.3.2 (1) dimensionality and metrics 318
  §9.3.2 (2) specifying a form of processing
      based on stress 321
  §9.3.2 (3) specifying a form of processing
      based on unit circles 335
  §9.3.2 (4) developmental trends seen
      in processings 339
  §9.3.3 conclusion 344

Chapter 10 translatability experiments - developmental
trends

§10.0 introduction 348
§10.1 experiment 1 348
  §10.1.1 experimental design 348
  §10.1.2 administration 350
  §10.1.3 hypotheses 351
  §10.1.4 results and discussion 354
§10.2 experiment 2 (developmental trends)
  §10.2.1 aim and hypotheses 381
  §10.2.2 results and discussion 388

Chapter 11 conclusion 396

Footnotes 400

References 403
Appendices I

Appendix 1: translation of Japanese sentences tested (see chap. 4) 445
Appendix 2: the analysis of nine TL sentences 454
Appendix 3: the basic distinctions of Japanese verbs of destruction (see chap. 4) 463
Appendix 4: examples from school textbooks
-- an examination of 'sure' items 464
Appendix 5: the definitions of inductive and projective limits 481

Appendices II Tables and Figures 489

List of Abbreviations

NL Native Language
TL Target Language
IL Interlanguage
L1 First language
L2 Second language

KOC3 KOwasu - Chuugaku 3-nen: 3rd year junior high school group
KOK1 KOwasu - Koukou 1-nen: 1st year senior high school group
KOK2 KOwasu - Koukou 2-nen: 2nd year senior high school group
KOUI KOwasu - University 1-nen: 1st year university group
KOUI2 KOwasu - University 2-nen: 2nd year university group
YAC3 YAburu - Chuugaku 3-nen: 3rd year junior high school group
YAK1 YAburu - Kukou 1-nen: 1st year senior high school group
YAK2 YAburu - Koukou 2-nen: 2nd year senior high school group
YAU1  YAburu - University 1-nen: 1st year University group
YAU2  YAburu - University 2-nen: 2nd year University group

MDS  Multi-Dimensional Scaling (here it refers to the nonmetric MDS, the MINISSA programme)

DIM or dim  dimension(s)
Tg  category boundaries
SV  Scale Values
DD  Discriminal Dispersion
VO  Variable Order
CO  Constant Order
N(0,1)  Normal distribution with mean (0) and standard deviation (1).

Cov  Covariance
E  Expectation
\[ \Pi_{i \leq 1} b_i \]  Reads product of \( b_1, b_2, \ldots, b_i \) : \( b_1 \cdot b_2 \cdot \ldots \cdot b_i \)

\[ \sum_{i \leq 1} b_i \]  Reads sum of \( b_1, b_2, \ldots, b_i \) : \( b_1 + b_2 + \ldots + b_i \)

\( \tau \) or \( r_s \)  Kendall's tau
\( \chi^2 \)  chi-squared
W  Spearman's rank order correlation coefficient
p  Kendall coefficient of concordance

SD  Standard Deviation
M  Mean

F-matrix  Frequency matrix
P-matrix  Cumulative frequency matrix
X-matrix  The matrix of normal deviations

lub  the least upper bound
glb  the greatest lower bound
MOD  MODular lattices
DIS  DIStributive lattices

Sig.  Significant
N.S.  Non-Significant
Chapter 1 Introduction

§ 1.0

This thesis is concerned with the question of how we can explain the judgement of translatability made by Japanese learners of English. Kellerman (1977 & 1978) has shown that the investigation into the translatability judgement is useful to the study of the process of language learning and acquisition. The present thesis is in this sense a follow-up study of Kellerman's research.

I have modified Kellerman's study in six ways:

(1) Kellerman hypothesized that a learner's 'feel' of translatability would be based upon the extent of "coreness" reflecting one of the "language independent" universals. He used the multi-dimensional scaling (MDS) to represent the NL inter-lexical structure. He then interpreted one of the dimensions as "core dimension". This interpretation is "putative" (Kellerman, 1979), but the data showed high correlation with the core dimension. The present research is not concerned with this 'core' hypothesis. Instead, I hypothesize there is a 'sure' item whose translatability a learner is confident of. (The syllabus by which a learner was taught would indicate the most probable 'sure' item. In some experiments, the subjects were supplied with an experimental 'sure' item.) I hypothesize that a learner's translatability judgement is based upon his 'feel' of similarity between his 'sure' NL item and the other items he is asked to judge for translatability. We assume that he uses his 'sure' item as the basis of his decision-making. The present assumption is more realistic than Kellerman's core hypothesis, since we regard the judgement of translatability as the interaction of syllabus and a learner's use of analogy (i.e., to make use of his 'feel' of similarity between a NL concept and a TL word).

(2) I make use of the MDS distances rather than the MDS dimensions (in one pilot study). The learner's 'feel' of similarity among NL items is measurable. MDS computes...
distances between all possible paired items. These derived distances reflect the degree of similarity between paired items. In the present situation, they are indexes for the similarity relationships between NL items. The relationship a learner would intuitively conceive may be best regarded as some vague similarity relationship (i.e., he cannot explicitly analyse what the relationship is.) In this sense, his translatability judgement seems to involve analogical process. Because MDS computes distances between all pairs of NL items, we could work out how a learner has arrived at his conclusion of translatability, by looking at the derived distances and the judgement data. However, as we see below, the MDS distances are not relevant in predicting the judgement of translatability.

(3) I have carefully controlled the experimental items, using well-known psychological methods (see Chap. 5 & 6). This allowed the present research to take advantage of other psychological methods such as free recall (Chap. 6) and psychometric experiments to attest the accuracy of measurement (Chap. 7). I have also made use of the Thurstone-Torgerson method to estimate the 'feel' of similarity between paired sentences.

(4) I have tried to clarify, following Rosch (1976), what we mean when we say 'word-meaning' (Chap. 3). This tentative definition of word-meaning provides the basic framework of analysis shown below (Chap. 9 & 10).

(5) I have incorporated several kinds of data-processing into hypotheses (chap. 4, 9 & 10). Corder (1976) often indicates the importance of data processing, hypothesis formation, and hypothesis-testing in language acquisition (see, §1.0). The main interest of the thesis is in the effect of data processing upon the judgement of translatability. Several kinds of information processing were introduced into the hypotheses (Chap. 4, 9 & 10). Of the 6 types of processing considered here (chap. 9), "categorical processing" turned out to be the most important of all. The definition of categorical processing is based upon Shepard (1974) and accords with the present definition of word-meaning.

(6) I have tried to develop a model of the relationship
between the method of processing and lexical structure. As Garner (1976a & 1976b etc.) suggests, the subjects adopt a form of processing that corresponds with the subjective structural understanding—that is, they process as they understand the structure. According to this view, structural representation should have one-to-one correspondence with processing. The experiments in chap.9 suggest that all five groups of subjects adopted "categorical processing" when making similarity judgements. Shepard (1974) argues that the categorical structure is not suitable for dimensional scaling. This suggests that the multidimensional scaling adopted by Kellerman is not a valid technique for linguistic materials, whereas the Thurstone-Torgerson method can encompass the judgemental limits of translatability.

Section 1.1 introduces the basic terms which are closely related to the research topics and methodologies in interlanguage (IL) studies. The last two quotations in §1.1 are particularly relevant to the present research. This summarization may not be necessary; however, there are many different views on what IL studies are, and for this reason, I felt it useful to present the background of the terms I will be using in my research. Section §1.2 describes how the thesis is organized.

§ 1.1 basic terms in IL studies

The term 'interlanguage' was first used by Selinker (1972). It is a collective name representing the language produced by second language learners or acquirers. Such languages are only partially well-formed in terms of the rules of the target language and they usually include approximate (intermediate or deviant) forms. It is assumed that going through overlapping and non-discrete stages of development a learner's personal 'transitional competence' gradually approaches the system of the target language. This developing competence is "dynamic" and "goal-oriented" (Corder, 1977b: 1981: 90). For this reason, Nemser first offers the term 'approximative system' for the description
of a learner's language (Nemser, 1971).

The study of interlanguage is the study of the language system used by language learners. A language learner is assumed to possess his version of 'a set of rules' with respect to his target language (Corder, 1976). A learner's performance, whether erroneous or well-formed in terms of the real rules of the target language, is seen as an indication that he is using a "definite system" of language at every point in his development (Corder, 1967). The learner's errors are viewed as evidence of the existence of this system (ibid). The errors are, in theory at least, systematic when they are generated by the definite system of interlanguage (ibid).

This learner-generated sequence is called a "built-in syllabus": the former is cited by Corder (1967) from Mager (1961); the latter is used by Corder (1967, 1978, & 1979). Corder quotes Mager saying that the information sequence that is "meaningful to the learner" is "assimilated by the learner". This statement could be interpreted as a forerunner of Krashen's notion of "comprehensible input": (Krashen & Terrell (1983) and Krashen (1985)). When the information sequence available to a learner allows for the natural development of cognition reflecting the internal state of the learner's grammar, it would facilitate learning and new input would be assimilated into his grammar.

Corder (1978c) adopts the Piagetian distinction between assimilation and accommodation (see Chap. 4). The process which brings forth some change in the cognitive structure is termed accommodation. In this sense, what Mager refers to as 'assimilation' could mean 'accommodation'. What is "comprehensible" to a learner when he faces any new experience is determined by how much sense he can make of the new experience. To what extent he can make sense out of it is determined by the knowledge he has gained from his previous experiences (Corder, 1976). For this reason, the sequencing of teaching materials is highly important to facilitate learning. The study of
interlanguage aims to assist in the improvement of teaching materials and methods (Corder, 1981), and in the development of a natural approach in language learning (see Krashen & Terrell, 1983).

Not all utterances made by a second language learner are erroneous. Some are well-formed, either consistently or by chance. A deviant form will not be permanently used by the interlanguage speaker. One noteworthy characteristic of an interlanguage speaker is a greater degree of variability than that observed among native speakers. The interlanguage performance varies depending on the intra-sentential environment, inter-sentential environment of discourse, extra-linguistic context of situations and extraneous matters of premature linguistic planning (e.g., a shift of topic in the middle of conversation). The last three factors are equally recognizable as sources of variable performance among native speakers as well. In this sense, the higher variability among second language learners is assumed to be a matter of degree. Once it is felt that any second language learner possesses "a more or less well defined personal grammar to base his utterances on" (Corder, 1976: 1981,66), "we can talk about his performance as being rule-governed in the same way as we speak of the performance of any native speaker of a language as rule-governed" (idem).

Corder uses the term 'idiosyncratic dialect' in reference to a learner's language, rather than 'idiolect' (Corder, 1973). This is because the latter term is normally applied to a native speaker's personal competence in which an institutional, stable social code is reflected (Block, 1948; Hall, 1951; Hockett, 1958; Ullman, 1972). Thus, a native speaker's idiolect is considered stable, but in a learner's interlanguage, one cannot expect such stability. Yet, the learner's variable IL can serve a social function:

"...the learner's 'knowledge of the language' ... is the 'model that the learner has of the target language'. The model is inaccurate in various respects,
Idiosyncratic dialects involve the following four kinds of deviant language use: (1) the figurative poetic use of language; (2) the speech of an aphasic; (3) a child's language; (4) a second language learner's language (op. cit.). Corder suggests three reasons for regarding these deviant languages as belonging to the same category "idiosyncratic dialect": (1) They all contain some deviation from target norms. (2) But, they are rule-governed, as long as we regard them as the product of a psychological entity called "grammar" which is stored in our memory and which is constantly at our disposal. (3) Furthermore, in actual use, these deviant dialects can carry messages about the psychological phenomena a language user wishes to convey.

In order to stress the parallels among the above 4 kinds of language use, the neutral term of 'dialect' has been chosen. It has been suggested that a deviant idiosyncratic dialect contains some normative system of TL. In this sense, these dialects are in part well-formed. There is at least something shared between the deviant dialects and their TL. The more a speaker of a deviant dialect learns and acquires the TL, the more the common parts increase and expand (see Diagram 1). This corresponds to the idea of transitional competence. Transitional competence describes the process in which the intersection in Diagram 1 broadens in the course of learning:

![Diagram 1](attachment://diagram1.png)
An IL speaker becomes aware of his insufficient linguistic resources, when his communicative needs are not met. This relationship between linguistic resources and communicative needs is also covered in IL studies, since it relates to problems in the classroom, i.e., how to create motivation to learn the TL. By bringing a learner's linguistic resources into line with the conventions of the TL, his IL system becomes an effective means of communication within the TL society. Since a learner's awareness of his linguistic limitations is sharpened by increasing his communicative needs, some interlanguage studies support a communicative method of language teaching (cf. Richards & Schmidt, 1983; Færch & Kasper, 1983, etc.)

We have seen above that the term built-in syllabus was used to refer to the learner-generated sequence. We have also seen that the assumption of a definite system in a learner's cognition can account for the logical development of the rule-governed learner's target language.

What gives rise to a learner's definite system in his cognition? In Chomsky's view of language acquisition a learner's personal competence is assumed to be the result of an inherent faculty of mind. Corder argues that there are obvious and undeniable differences between the acquisition of a first language and the learning of a second language, due to such external "circumstances" of learning as maturational differences, motivational differences and attitudes. It is these external and contingent factors that cause the apparent differences between first language acquisition and the learning of a second language. The differences become unimportant once we look at the internal processes which take place in the learning of a second language (Corder, 1979). These internal processes are regarded as the manifestation of "an innate predisposition to acquire language" (Corder, 1967: (1981: 7)) or "inherent properties of the human mind" (Corder, 1981:203). Whether a given language is a native language or a second language, the acquisitional process will not show any fundamental difference, as it follows the same innate process of
language acquisition.

It is important to note that Corder regards the built-in syllabus as offering an alternative to the Skinnerian view of errors (Corder, 1979). According to Skinner, a learner's errors are the result of a transference of the mother-tongue. After a brief discussion of the climate in which contrastive analysis was adopted as a popular method of predicting a learner's difficulties, Corder explains as follows:

"With post structuralist developments in linguistics, associated with the name of Chomsky, a willingness to seek common or even universal features in human languages became again a goal of linguistics, but now explicitly explained in psychological terms as inherent properties of the human mind. Language acquisition and second-language learning could now be approached as a problem of cognitive learning and the possession of knowledge of a certain kind ('competence') rather than as a set of dispositions to respond in a certain way to external stimuli. A language user possesses a set of cognitive structures acquired by some process of data-processing and hypothesis formation in which the making of errors was evidence of the learning process itself and probably not only inevitable but necessary (Dulay & Burt, 1974). It now became relevant to study a learner's linguistic performance in detail in order to infer from it the nature of that knowledge and the processes by which it was acquired. ...Central to the investigations was the analysis of the errors made by learners since they represent the most significant data on which a reconstruction of his knowledge of the target language could be made. This is essentially the point of view presented by Corder (1967)."

(Underlying added.) (Corder introduces the term 'built-in syllabus' for the first time in his article in 1967.)

Corder, (1979: 203)
The study of IL aims to reconstruct a learner's knowledge of his TL. His version of the TL is the outcome of his data-processing and what he "induces" as a rule of the TL (Corder, 1967). The process in which a learner induces a rule is likened to the formation of a hypothesis. Since his utterances reflect his induced rules, the IL speech has been called 'hypothesis-testing'. Thus, there are roughly three stages in the acquisition of a second language: (I) data-processing, (II) hypothesis-formation and (III) hypothesis testing.

Corder regards the processes of data processing and hypothesis forming as crucial in the acquisition of a second language. Since a learner does not comprehend all that is supplied as input, the mechanism of data processing is assumed to intervene between "input" and "intake". Corder argues that the plausibility of such a mechanism is recognizable when we note the difference between what has been made available to a learner and what is internally accessible to him (see Chap. 5 & 9).

As the above quotation indicates, the hypothesis of the built-in syllabus contains the following two related claims. (1) The acquisitional process of a second language is guided by cognitive principles. In the above, data processing and hypothesis forming are suggested as possible principles. Later, Corder includes simplification as another possible candidate. For this reason, Corder regards research into pidgin languages or free learners outside educational institutions as promising in the establishment of the natural processes of language-acquisition. (2) When these language-independent cognitive principles are actually in operation, there is some universal (natural) order in second language acquisition.

The second claim of the built-in syllabus has been experimentally examined by many investigators in the area of the acquisition of morphemes. Although some methodological weaknesses have been pointed out (see Chap. 2), these morpheme studies are the first research results
which confirm the plausibility of the built-in-syllabus hypothesis. Krashen (1981) and Dulay & Burt (1980) generalized these results into four stages of development. This part of the built-in syllabus was called the natural order hypothesis in the 1980s.

Previous major studies, naturalistic or experimental, have adopted performance analysis (i.e., the analysis of natural speech or elicited speech: see also §2.1). When analysing performance, the existence of the cognitive processes (i.e., the first claim of the built-in syllabus) comes out in the course of the discussion. However, as Felix (1985) mentions, this first claim of the built-in syllabus which he calls a "problem solving cognitive system" "is still largely unknown" (Felix; (1985:70)).

The present study aims to support the relevance of the first claim of the built-in syllabus hypothesis. For this reason, it was necessary to review some fundamental experiments concerning data-processing and hypothesis formation (concept-forming). Corder's following observation was taken as my guideline:

"...there is a property of the human mind which determines the way language users process data to which they are exposed, whatever the superficially different properties of the data may be."

(Underlying added.) Corder, (1981: 72)

The guiding principles suggested in the quotation are that one should look into previous experiments which have established that the hypotheses concerning data-processing and hypothesis forming are not influenced by the nature of materials. What conditions are necessary for us to claim that something is materially independent in nature? The following two points are suggested:

1. The phenomena to be reviewed must be valid for both children and adults in their first language.
2. The investigations must be carried out, using both linguistic and artificial materials.
Chap. 3 presents previous psychological experiments concerning concept-formation which satisfy the above two conditions. The discussion relates to some previous studies in lexical acquisition among first language children. The 6 kinds of data processing reviewed in chap. 9 partially satisfy the first of the two points mentioned above, since the subjects were university students in the department of psychology, and there were some experiments among L1 children in which the hypotheses proposed by Garner are incorporated. In this sense, the literature looked at in chap. 9 meets the suggested guidelines.

§ 1.2 Thesis outline

This research is concerned with the question of how we can explain the judgemental process of translatability, and the issue of whether or not research into the translatability judgements can illuminate learning processes, as Kellerman's experiments have suggested (1977, 1978 and 1979). The experiments deal mainly with two polysemous Japanese verbs 'kowasu' (break) and 'yaburu' (tear), involving five groups of Japanese learners of English. The experimental methods are intended to be "learner-oriented" and "process-oriented", in accordance with a research aim of inter-language. The experiments reported below try to demonstrate that the judgemental process is best seen as an interaction between data-processing and the analogical process. Both data-processing and the analogical process are known to be an important part of language learning or acquisition. For this reason, the present research provides an explanation not only for translatability judgements but also learning (or acquisitional) processes. Furthermore, the present interactive view of the judgemental process accords with Corder's view of language acquisition which consists of the three stages of data-processing, hypothesis-formation and hypothesis-testing (see above). As we see below, the judgemental process of translatability involves all these stages.
There are seven experimental studies in the thesis. Since the experimental work starts with a follow-up study of Kellerman (1977, 1978 & 1979), Chap.2 gives a selective review of the literature in relation to Kellerman's study. Since each chapter gives an overview in its respective field, the review in chap. 2 is selective and aims to provides the background with which Kellerman's study has been associated (see also §4.1).

Chap. 3 gives a tentative definition of word-meaning and makes an operational distinction between competence and performance. According to Rosch's "prototype" hypothesis (see Rosch, 1975, Mervis and Pani, 1980, etc.), a language-learner needs to grasp not only a prototype (which can be the "best exemplar" derived from the "basic-level object", or simply a poor exemplar), but also the limit of application of each lexeme. The proponents of the "prototype hypothesis" have also demonstrated that a language-learner knows the "maximal within-category similarity" and the "minimal between-category similarity". These notional characterizations are not sufficiently explicit. This chapter offers some tentative definitions to these notional characterizations. Boolean algebras can show that the prototype which is the best exemplar at the basic level stands for the greatest lower bound (glb). Furthermore, the maximal within-category similarity and the minimal between-category similarity defines the limit of application of each lexeme; and the former corresponds to the glb, and the latter, the least upper bound (lub). By examining four classes of over-extensions among L1 children, we see that a simple criterion of all vs some features can distinguish competence from performance. The criterion of all relevant features corresponds to the prototype at the basic level which has the "most features in common" and which we have represented as the lub. The representation by the lub ties up with Shapard's characterization of "categorical structure" which he posited as the relevant representation of linguistic materials (1974). Further, the representation of lexical performance by means of the lub accords with Tversky's
view of the comprehension process of metaphorical expressions (1977). Moreover, the present distinction between competence and performance accounts for Ullmann's characterization of word-meaning (1972): i.e., the "reversible and reciprocal relationship between sound and sense". The last point is explained by Piaget's INRC which is a class of Boolean algebras and complemented distributive lattices. The present definition of word-structure is also suitable for characterizing five types of processings (separable, asymmetric separable, dominant, integral and categorical processings: see chap. 9 & 10).

Chap. 4 presents pilot study (1) which is my follow-up study of Kellerman's experiments. But the follow-up study includes some modifications of his research: viz. (a) the effect of processing upon judgement data; (b) the effect of the 'sure' item in the analogical process; (c) the distinction of 'consistent' vs 'variable' responses in correspondence with the interlanguage distinction between competence and performance. The same judgement data were analysed in three ways: Analysis I, II & III. Each analysis adopts a specific processing assumption, and we try to examine which assumption provides the best explanation for the data obtained. We consider three kinds of processings: "asymmetric separable processing, separable processing and integral processing". Garner (1974). Analysis I replicates Kellerman's implicit assumption of 'asymmetric separable processing': i.e., Kellerman interprets his data as suggesting that one-dimension which he calls "core/non-core" dimension plays a major role in the judgemental process; this interpretation means that he has implicitly assumed asymmetric separable processing on the part of his subjects. Thus, Analysis I adopts this asymmetric separable assumption. In Analysis II, the asymmetric separable assumption is contrasted with the integral assumption. It also includes another important consideration: i.e., we consider the effect of 'sure' item in the judgemental process. We assume here that there is a 'sure' item whose translatability the subjects are confident of. That is, we regard the
judgement of translatability as based upon a learner's 'feel' of similarity between his 'sure' item and a given item he is currently judging. The 'sure' items can be either taught or are erroneous IL forms which have been fossilized as 'sure' in his acquisitional process. For this reason, the major school textbooks the subjects used were consulted (see appendix 2). Some provision is made to accommodate the latter type of interlanguage 'sure' items within the present criteria of 'sure' items: see §4.3.3. Analysis III adopts the integral assumption, along with the assumption of 'sure' items. The main feature of Analysis III is that the MDS measurement of similarity is contrasted with the measurement of similarity by the Thurstone-Torgerson method. The judgement of translatability relates to an analogical process which is revealed in a learner's reliance on his 'feel' of similarity. Both the MDS distances and the distances derived from the Thurstone-Torgerson method measure a learner's 'feel' of similarity. It is interesting to see which method of measurement, the MDS method or the Thurstone-Torgerson method, can account for the judgement data better.

This pilot study (1) deals with four verbs: kowasu (break), yaburu (tear), kudaku (smash) and oru (bend). The group of subjects participating was the most advanced group of all. Analysis I reveals (1) that kowasu and kudaku justify Kellerman's implicit assumption of 'asymmetric separable processing', while the other two verbs do not. (2) While kudaku supports Kellerman's interpretation of "core/non-core" dimension, the other three verbs do not justify his interpretation. Analysis II provides evidence (1) that in the judgemental process both NL assimilation and TL accommodation occur jointly. This is revealed in two ways. First, while one NL dimension correlates positively with the judgement data, the other NL dimension correlates negatively. Second, this negative and positive contribution of NL is reflected in the predictive capacity of integral assumption which turned out to be comparable to that of the asymmetric separable assumption. (2) The
analysis also indicates that the subjects categorize the items on the NL lists in making the translatability judgements. This effect of categorization is further tested by the two free recall experiments in chap. 6. (3) The data also suggests some evidence of what is called here 'concretization' (transforming the initial abstract input into concrete). This appears to be initiated by the syllabus by which the subjects were taught and the feature of the syllabus is explained in relation to the "dominance effects" in psychology. Analysis III indicates that the distance measures derived from the Thurstone-Torgerson method are comparable to the MDS distances in explaining the analogical process. The implication of this finding is discussed in §4.4.3 and §7.0. This follow-up study indicates clearly that it is necessary to make our hypotheses more specific, by running further pilot studies to investigate specific topics separately (see §4.6).

Pilot study (2) in chap. 5 is a preliminary to the following five experimental studies in chap. 6-10. The experiments in this study are to determine (1) object nouns to be used as experimental items, (2) the number of steps in the rating scale and (3) the method of elicitation of input data for MDS. The object nouns were chosen by standardizing the "conjoint frequency effects" as a measure of "objective familiarity". It was decided to make the number of steps of the rating scale six. It was also decided that we should elicit the MDS input data from the subjects by the method of 'categorical sorting'.

Chap. 6 presents Pilot Study (3) involving two free recall experiments. This pilot study has two purposes (see §6.0). (1) We examine whether the clusters produced by MDS are psychologically real or not. (2) We also examine whether the translatability judgement test encourages the subjects to categorize the items which are semantically related. This is to attest the categorization effect speculated in chap. 4.
Prior to the free recall experiment, the object nouns elicited in pilot study (2) are examined in terms of "subjective familiarity". Since the objective familiarity is standardized by pilot study (2), the two areas of well-known biasing factors mentioned above are kept constant in the materials used below.

MDS produces a map in which items are grouped together (i.e., configuration). It is known in psychology that people recall items in semantically related groups (i.e., "subjective organization"). We try to see whether the clusters produced by MDS correspond to these subjectively organized clusters in free recall results. A fair correspondence would offer some evidence for the MDS clusters. Since the MDS results are essentially the outcome of mathematical operations, the method of free recall is a meaningful way of proving some psychological reality in the MDS configuration. The experiments yield some affirmative evidence in this respect.

Two groups of subjects participated in the free recall experiments. The experimental group was incidentally asked to recall items after the translatability judgement test. The control group performed only free recall. The number of clusters was counted on the basis of Monk (1976). The results showed that the experimental group produced significantly more clusters in free recall than the control group did. This indicates that the translatability test encouraged the subjects to categorise the items on the NL lists. The greatest number of clusters were 2-item groupings. This suggests that the subjects compared two items at a time in making their judgements. This finding led me to adopt the equation of paired comparison known as the Thurstone-Torgerson method discussed mainly in chap. 7.

Chap. 7 presents Pilot Study (4). It examines the Thurstone-Torgerson method (SIM), following the experimental rationale by Atkinson and Juola (1974). There are two experiments. Experiment 1 examines whether the scale
values derived from Thurstone's (Torgerson's) equations offer reasonable measures in relation to the category boundary. Experiment 2 examines whether the predictions about the response times derived from the equations correspond to the observed response times during which the subjects compare pairs of items. In both cases, the extent to which the predictions correspond to the experimentally observed values determines the degree in which the Thurstone-Torgerson method is adequate for the purpose of the translatability judgement process. The two experiments justify our application. The experimental testing of the method not only guarantees the accuracy of SIMs in psychological terms, but also, guarantees the usefulness of SIM for estimating judgemental limits (inductive limit as well as projective limit).

The essential part of the judgemental process of translatability relates to the analogical process. In this sense, the measurement of similarity among NL items is crucial. Since the smallest analogical process is two-item based, the Thurstone-Torgerson method which is derived from a paired comparison task is adopted here. The above experiments in chap. 7 show that the Thurstone-Torgerson method is useful for estimating judgemental limits (inductive limit as well as projective limit). The former term is applied to the cases in which a learner induces the limit of application of a novel lexeme at the initial exposure to a TL lexeme. The latter term is used for the cases in which a learner uses his already acquired judgemental limit to the other NL words, in judging their translatability. The free recall experiments in chap. 6 also indicate that the present application of the Thurstone-Torgerson method is adequate, since the subjects have produced a significantly larger number of 2-item groupings during their free recall.

§7.1 argues that the Thurstone-Torgerson method is a simpler way of evaluating the bivariate normal distribution function. According to this interpretation, the present way of evaluating the inductive or projective limit
satisfies the conditions by which these terms are explicitly defined by Takeuti (1978): see Appendix 5. In this sense, our use of the terms, inductive or projective limit, is mathematically justified as well.

We have seen in §1.1. that language acquisition involves the three stages of data-processing, hypothesis-forming and hypothesis-testing. It is important to bear in mind that the acquisition of inductive limit is a crucial part of hypothesis-formation (see chap. 3 & §7.4). The prototype theory of lexical acquisition claims that a language-learner needs to know the limit of application of any linguistic rule including a lexical field. This limit of application of a linguistic rule is called here an 'inductive limit', when it is induced for the first time. Likewise, the projective limit plays an important role at the stage of hypothesis-testing in which the acquired judgementsal limit is actually used at the time of utterance. In this sense, chap. 7 demonstrates how the notions of inductive and projective limits can be mathematically realized so that we have a means of illuminating the cognitive functions of inductive and projective limits experimentally (see also chap. 10).

Pilot Study (5) in chap.8 takes up the hypothesis of compound vs coordinate bilinguals. Compound bilinguals possess a single language system, while the coordinate bilinguals possess two separate systems, one for NL and the other for TL. If the coordinate language theory holds true, the present research (concerning the translatability judgement) would not make sense, because if the two language systems are disconnected, the subjects would not be able to make any translatability judgements. The purpose of pilot study (5) is to check whether the subjects participating were completely coordinate bilinguals or not. The data is analysed in three ways: the Wilcoxon matched-pairs signed-ranks test, the pairwise t-test and the probability ellipses. The analyses show that the subjects participating do not possess the extreme form of a co-ordinate language system.
Chap. 9 presents Pilot Study (6). It deals with the question of what kind of information processing the subjects adopted at the time of making similarity judgements. For this purpose, MDS was used. According to Garner (1974) and (1974^), there are clear connections between metrics and processings. One can determine a specific form of processing by adjusting metrics in relation to "stress" (i.e., the index to reveal the amount of discrepancy between input and output). In this pilot study, all five groups of subjects who took part in the translatability tests in chap. 10 were tested. The five groups tested on the two verbs (yaburu & kowasu) appear to have processed the NL interlexical structure categorically, adopting categorical processing. As Shepard (1974) indicates, when a set of experimental items is categorical, a dimensional scaling of MDS is not appropriate. This suggests that the MDS distances would not be useful for the purpose of predicting translatability judgements.

Chap. 9 considers six types of processing: configural, separable, asymmetric separable, integral, dominant and categorical processings. The explication of categorical processing starts with Shepard's definition (1974). By relating his definition to the present definition of lexical structure (chap. 3), Boolean algebras clarify that categorical processing requires the parallel operation of the other four processings. The discussion shows that categorical processing relates to the present definition of competence and that a specific use of the other processings, when used separately, reflects such performance factors as the nature of the experimental task and the nature of the experimental materials. The present experimental results indicate that all groups of subjects have essentially adopted categorical processing. The examination of 'unit circles' and 'stress' shows at the same time that apart from one group (KOU2) involving kowasu the other groups adopted dominant processing in coping with their performance demands.
Chap.10 presents the results of the translatability tests among the five groups of subjects involving the two Japanese verbs kowasu and yaburu. The SIM distances, (instead of the MDS distances,) were used to estimate the 'feel' of similarity between the 'sure' items whose translatability the subjects were confident of and the items the subjects were judging for translatability.

The translatability experiments in §10.1 simulate the syllabus design by which the subjects were taught. The syllabus is explained in chap. 4 & 6 in relation to the "dominance effects" in psychology. (1) The results (§10.1.2) support the dominance effect and Rosch's hypothesis of "basic level" (see also appendix 4). (2) The notion of inductive and projective limits which were introduced in chap. 1, 3 & 7 is also justified by the data. The inductive and projective limits fall in the notional range of 'some relationship' ( Category boundaries $T_2$ - $T_3$ ). (3) The 'sure' items ( definition is given in chap. 4) are those whose translatability the subjects are confident of and on which they base their judgement of translatability. Some 'sure' items were taught at school but the other 'sure' items were not taught but self-generated. The latter self-generated 'sure' items are examined in view of Boolean algebra. The discussion shows that the self-generated 'sure' items are all characterisable as "relative complements" of an experimentally instructed 'sure' item. The discussion suggests that the self-generated 'sure' items are derived from 'the principles of maximal similarity and identity'. (4) The developmental trends among the five groups of subjects are discussed in §10.2 in view of (a) the effects of processings upon the translatability judgements, (b) concretization vs abstractization (c) differentiation vs unification and (d) the U-shaped behavioural growth. It appears that all these phenomena can be explained within the framework of Boolean algebras.

Chap.11 gives some tentative conclusions.
The analogical process is quantified in terms of the Thurstone-Torgerson method, but the method is not adequate to reveal the identity principle for which Matsumoto's theorem is used (cf. §10.1.4 (4)). But the Thurstone-Torgerson method is useful for the principles of semantic closeness and maximal similarity.
Parallel between lexical acquisition and the judgemental process of translatability

Data-processing
Chap. 9

Hypothesis-formation
A language-learner
induces a judgemental limit
(inductive limit)
as well as a prototype, but
in the case of the translatibility
judgement, the induction of
prototype may be substituted
by his NL knowledge: cf. chap.8.

Hypothesis-testing
A language-learner
tries out his acquired
judgemental limit
(projective limit).

chap. 3 (model)
chap. 7 (the method of computation)
Categorization (§4.4.2 Analysis II, §4.4.3 Analysis III, §6.3.0 and §6.3.3) relates to the lexical reorganization discussed in terms of the three principles of similarity (cf. §10.1.4 (4)).

Prototype: chap. 3 and appendix 4.

Basic level objects: chap. 3, appendix 4, §10.1.3 (1) and §10.1.4 (1).

Chained reasoning: §3.4, §10.1.3 (3) and §10.1.4 (3).

Inductive and projective limits: chap. 3, chap. 7, appendix 5, §10.1.3 (2) and §10.1.4 (2).

Effects of processing upon translatability judgements: chap. 4 and chap. 10 (particularly §10.1.3 (4), §10.1.4 (4), §10.2.1 hypothesis 2, and §10.2.2 (result 2)).

Concretization: §4.4.2 (Analysis II), §10.1.1, §10.1.4 (1), §10.2.1 (4), §10.2.2 (result 4).

Dominance effects: §4.4.2 (Analysis II), §6.3.0, §6.3.3, §6.3.4, §10.1.1, §10.1.3 (1) and §10.1.4 (1).

Unification vs Differentiation: §9.3.2 (4), §10.1.4 (4), §10.2.1 (1) and §10.2.2 (result 2).

U-shaped behavioural growth: §10.2.1 (3) and §10.2.2 (3).
Chapter 2. Selective review of literature

§ 2.0

This chapter presents a selective review of the literature in relation to Kellerman's study (1977, 1978 & 1979). Since each chapter outlines an overview in its respective field, this review is brief and selective. The experimental method adopted by Kellerman is explained in chap. 4. I regard his method as "learner-oriented" and "process-oriented" (and a reason for this is given in chap. 4). On the other hand, as Long and Sato (1984) characterize, the methodology adopted by the proponents of the "natural order hypothesis" stands for "performance analysis".

Firstly this chapter reviews the major studies related to the "natural order hypothesis", with special emphasis on the methodology. The review will indicate that the learner-oriented and process-oriented methodology is an important development of the "performance analysis". Secondly, I refer to the view that lexical simplification, overgeneralization has been regarded as involving analogical processes; in fact overgeneralization errors are sometimes called analogy errors (e.g., Dommergues and Lane, 1976). The brief reviewing is necessary because of the nature of present experiments. In one of the experiments of translatability (chap. 10), the youngest subjects were taught that one of the experimental items on the NL list is translatable into English. They were asked to judge the translatability of the remaining NL items. In the case of older subjects (chap. 4 and 10), we know in advance which NL items they learned at school. In these cases, the above initial instruction was not supplied to them. We can expect that these experimental designs would elicit many analogy errors. For this reason, the brief review of lexical simplification is presented below (§2.2). In the course of discussion, I touch on the point that the distinction between 'recreation and restructuring hypotheses' relates to Piagetian notion of assimilation and accommodation and to Ausubel's theory of assimilation.
The well-known natural order hypothesis started with a series of experiments run by Dulay and Burt (1973 and 1974a-d). The experiments are also called "morpheme studies". They adopted the same research method used by the researchers in L1 language development (Brown, 1973, de Villers & de Villers, 1973, Cazden, 1972 and Dale, 1976). Showing the similar developmental pattern as that in the first language constitutes part of the argument to claim that the sequence obtained experimentally is 'natural' and innate.

The first study (Dulay & Burt, 1973) dealt with eight morphemes, involving the three groups of Spanish speaking children. The method of elicitation was a "structured conversation technique" later known as "Bilingual Syntax Measure" (BSM). (The BSM consists of seven or twelve coloured cartoons, depending on the versions (see Rosansky, 1976 and Porter, 1977). The BSM elicits twenty two up to forty responses from the child. It takes 10-15 minutes to complete the task. The experimenters write down the responses on the BSM answer sheets and the responses are usually tape-recorded. The experimenters make note of non-verbal gestures. Previous to the main task, there is some vocabulary comprehension check.) "BSM only looks at the degree of proficiency with which the child uses the structures he offers in response to the questions" (Dulay & Burt, (1974:48)). As in Brown (1973), they regarded each obligatory occasion as a test item. No occurrence of a morpheme in the obligatory context was counted as zero score, erroneous occurrence, as 0.5 and correct response, as 1.0. The rank order correlations based upon the average scores showed that the acquisitional sequences obtained from the three groups were highly similar.

Dulay and Burt (1974a-d) undertook a more extensive study. They dealt with eleven morphemes and compared Chinese
children with Spanish children. As in the previous study Dulay and Burt treated each obligatory occasion as a test item. They gave zero score for the non-occurrence of a morpheme in each obligatory context, one point for erroneous occurrences and two points for correct responses. This time, they used "group score method". The method is similar to the calculation of percentage. The denominator is the sum of all obligatory occasions and the numerator, the sum of the scores for each morpheme across all children. The ratio is multiplied by 100. "The acquisition sequences were obtained by ranking the morphemes according to decreasing mean functor scores." (1974:56).

They also adopted the "second language acquisition index" (SAI) based upon de Villiers and de Villiers (1973). Dulay and Burt regarded the 90 % criterion as a fair indication of aquisition. The SAI is then calculated as follows:

"The syntax acquisition index is the quotient resulting from computing a ratio whose numerator is the sum of the values of all the utterances of the child and whose denominator is the sum of the values of all the corresponding grammatical forms, multiplied by 100."

(Dulay & Burt, 1974: 59).

The indeces were divided into 5-point ranges. Each morpheme was assigned the value of the SAI ranges. Spearman's rank order correlation was calculated on the basis of the ranked data. The rankings were regarded as indicating the acquisitional sequence of 11 morphemes.

The ranking obtained by the methods of "group score", "group means" and SAI were compared again in terms of Spearman's correlation coefficients. The analysis showed that the three methods indicated the significantly similar rankings. The results were interpreted as

"providing a strong indication that universal cognitive mechanisms are the basis for the child's organization of a target language, and that it
is the \( L_2 \) system, rather than the \( L_1 \) system that guides the acquisition process".

(underling added.) Dulay, Burt, (1974\textsuperscript{d}: 65)

Dulay and Burt generalize the above view as "recreation hypothesis" or "creative construction hypothesis" which claims that the "restructuring" of the NL system is minimum in the \( L_2 \) acquisitional process.

Dulay and Burt (1975) ran the same experiment among the new Spanish group. They obtained the highly similar result to the previous two studies reviewed above. The acquisitional sequence was presented as consisting of four stages, making a special point that morphemes are not acquired one at a time, but that one group of morphemes as a whole is likely to be acquired earlier than the other groups. Dulay and Burt called this analysis "hierarchical analysis."

Fathman (1975\textsuperscript{a}), using the SLOPE test (Second Language Oral Production Test), obtained the supporting evidence for the above morpheme studies among Korean and Spanish students aged 6-14 years. Rosansky (1976) reports a 10-month longitudinal study involving two Spanish speaking children. She states that "statistically significant correlations were found between BSM-generated and spontaneous speech-based rank orders" (1976:423), but she indicates at the same time that because of the large variability in the \( L_2 \) learner's data, the comparison with the \( L_1 \) data such as de Villier's cannot be made in a meaningful way. She also indicates that the cross-sectional and longitudinal results for the same individual do not necessarily match, suggesting that the method of groups scores and group means might cancel out the real individual differences in the data. Hatch (1978) reports the case study which proved the point Rosansky made. Hakuta (1974) undertook a longitudinal study of a Japanese girl learning English who showed a poor correlation with the established natural order of acquisition.

Bailey, Madden and Krashen (1974) investigated the
possibility that adults might show a similar acquisitional sequence that was found among L₂ children. 73 adult subjects participated in the experiment. They were from the 12 different language backgrounds. By adopting the BSM, the experimenters found that the adult results agreed with those obtained by Dulay and Burt (1973).

Several other studies have investigated this possibility. Among others, Krashen, Madden & Bailey (1975) showed that a group of morphemes (rather than the individual morphemes being acquired sequentially) tended to be acquired earlier than the others, supporting Dulay and Burt's "acquisition hierarchy" (1975). Freeman (1975) dealt with 24 adults representing four different language backgrounds. She found "a high level of concordance" across the four groups, although "there was individual and language group variability apparent." Krashen et. al (1976) used the SLOPE test among the sixty-six adults representing five different language backgrounds. They found that the difficulty order obtained was not different from the order found among L₂ children by Fathman (1975). There was no significant difference between the adult speakers of different languages. These findings were interpreted as indicating that "certain similarities exist in the language acquisition processes utilized by children and adults" (1976: 145).

Krashen et. al (1977) elicited spontaneous speech from 33 adults from six different language backgrounds. The acquisitional sequence observed was close to the four-staged hierarchy reported by Dulay and Burt (1975). Krashen (1981) later reviewed the literature and indicated that a "proposed 'natural' order for second language" agreed with that observed in "Broca's aphasia" (1981:59). He regards as adequate the claim that there is an acquisitional order to which ESL learners adhere, in spite of their different ages and language backgrounds.

We have noted in reference to Dulay and Burt (1974) that the proposed 'natural order' has been regarded as the reflection of underlying cognitive activity (see also chap.1). But several studies investigated the possible influence
of other factors. Perkins and Freeman (1975) used two kinds of subjects, informal L₂ learners and formal learners, in order to find out whether the formal instruction can distort the natural 'order' of morpheme acquisition. They found that the formal instruction did not alter the order of acquisition, although the instruction improved the learners' performance. They also pointed out that "a rank ordering type of statistical analysis is obviously inadequate."

Larsen-Freeman (1976) investigated a possibility that the frequency of occurrence of the morphemes in adult native-speaker speech might have some correlation with the morpheme 'difficulty order' of L₂ children. She found a significant correlation.

Wagner-Gough and Hatch (1975) also indicate that the frequency of forms in speech the learner hears influences the language he produces. They argue that although there is some systematicity in L₂ acquisition, the appearance of forms in a learner's speech does not always mean he knows their functions. Further, they argue that the input the young language learner receives is different from the input the adult learner hears. Their study suggests that the method of counting the overtly appeared morphemes is not satisfactory to claim the authentic level of acquisition which includes the functional mastery of forms. It is even more unsatisfactory to claim some underlying cognitive mechanism, since the speech output the learner produces is the result of various input conditions under which he acquires his TL. It is important to bear in mind that the appearance of forms does not guarantee the mastery of their functions with which we normally associate with the acquisition of language.

Kleinmann (1977) investigated "avoidance behaviour" in adult L₂ acquisition. His study has an implication to the morpheme studies. As Perkins and Freeman (1975) states, on eliciting the data they came across instances of avoidance strategy: "our subjects were very adept at avoiding using the morphemes we were interested in" (1975: 239). Kleinmann (1977) found that avoidance behaviours were predicted by contrastive analysis and that since "affective" state of the learner correlated with the fre-
quency of the learner's use of TL structures, this psychological factor determined the amount of avoidance. His study is important in four respects.

(1) The morpheme studies take the face values of 'appearing' forms, but the learner may be deliberately avoiding forms. We have to contrive some means to force the learner to exhibit his competence as a whole.

(2) The strategy of avoidance is only one of the several communication strategies (see Váradi, 1983, Tarone, 1981, Færch & Kasper, 1983, etc). The other strategies may prevent the learner from using a certain morpheme, even though he knows its function and usage.

(3) If avoidance behaviour is brought about by the "affective" state of a learner, as Kleinmann's study indicates, the experimenter needs to know how much psychological factors determine the learner's production of speech. Or, we must be able to elicit the entire IL competence the learner possesses. Otherwise, as Hatch (1978) indicates, it is not appropriate to talk about the 'universal' order of acquisition on the basis of fragments of IL speech we elicit.

(4) Kleinmann's study shows that the contrastive analysis predicts the difficulty order and that the predicted difficulty order accorded with the amount of avoidance behaviours. In this sense, the occurrence of certain forms is governed by the learner's native language. This suggests that the production of speech relates to the learner's native language in this sense. This is contrary to Dulay and Burt's creative construction hypothesis discussed above.

Heilenman (1981) reports that the acquisition of morphemes relates to the learner's specific experience with a language rather than to his cognitive maturation. This study also suggests that it is important to take into accounts "extraneous" factors including the learner's experience with the TL, before we can talk about the 'underlying cognitive mechanism'. The global examination of the learner's TL input is a highly difficult problem for performance analysists. This necessity of entire examination of the learner's linguistic life was indicated by Corder (1981):
"A learner brings to the classroom many characteristics which are relevant to predictions about his career as language learner. These characteristics are the product of his membership of a community; he shares its language and its attitudes to, beliefs about, motivations for, and traditions in, language learning in general, and in the learning of specific second language. And he possesses particular features of personality as formed by his personal history of maturational and experience."

(Underlining added.) Corder, (1981: 95)

"...the 'value' to be assigned to 'well-formed' forms is only discoverable in terms of the whole system of his dialect. Thus, for example, a well-formed 'plural' or an apparently 'proper' use of the definite article can only be understood in relation to his 'ill-formed' plural or his use of other determiners.

This means that all the learner's sentences should in principle be analysed."

(Underlining added.) Corder, (1981: 21)

This methodological difficulty of performance analysis relates to the issue of well-attested method of elicitation. Porter (1977) found that the BSM tended to elicit some morphemes more often than the others: e.g., the morpheme representing past regular was not elicited even once (1977: 59). Moreover, when he used the BSM among L_1 children, he found that the L_1 BSM data did not correlate with the L_1 free speech data. On the basis of his findings such as these, he concluded that the previous 'natural order' of morpheme acquisition was probably an "artifact of the BSM".

Wode et. al (1978) indicates a shortcoming different from data collection procedures and statistics used (see Rosan- sky, 1976 and Anderson, 1977). On the basis of their longitudinal studies among four German children, Wode et. al
present five points which are overlooked by the morpheme studies: (1) the importance of overgeneralization (i.e., the child assigns two or more functions to one form); (2) the frequent occurrence of intermediate forms; (3) the importance of avoidance phenomena; (4) the learning strategy of "stereotypes" similar to Hakuta's "prefabricated chunk learning"; (5) the learner's reliance on L1. They also argue that 'mean values' based upon the cross-sectional studies tend to 'obliterate' the complex covert mental activities listed above.

Wode et. al also imply that the analysis of the morpheme studies which fail to take a learner's intention and point of view is target-oriented and misses the transitional stages of development which an individual learner goes through. Brown (1983) indicates the same point. Related to this, Hatch (1978c) views that the analysis obscures the developmental patterns because the data-treatment overlooks different contexts, functions and uses of a given morpheme. Pica (1983) indicates the same point. In this sense, the morpheme studies were not successful in achieving one of the research aims in IL studies, i.e., to describe the IL transitional competence of a "learner-language" (see chap.1). The notion of learner-language is offered by Corder for the researcher to adopt the viewpoint of a learner and to recognize the linguistic functions which the IL forms are intended to convey to us.

§ 2.2 learner-language and process-oriented research

There seems to be several promising research interests in the more recent IL studies which try to incorporate a learner's point of view. Among others, the following three trends are noteworthy:

(1) investigations into the language learner's intended syntactic functions in the context of discourse (e.g., Hatch, 1978b, Huebner, 1985);

(2) investigations into the learner's judgemental data (e.g., Gass, 1983a and Sorace, 1985);
(3) investigations into the possible sources of language transfer or into a learner's "psychotopology" (e.g., Kellerman, 1979, Gass, 1983 and Schachter, 1983).

The present research relates to the first half of the third approach. The merit of this approach is that although the investigator remains an outside observer who cannot go into the learner's black box, it is possible to make experimentally some internal process externally observable, as long as we can be confident of the point that language transfer is actually happening. This is because it is possible to represent the NL structure of some sort in a visual form by the learner-oriented method, as Kellerman's study illustrates. If the reproduced NL structure were a faithful representation of the internal structure, it would appear to be possible to even follow the learner's inference which brings about his decision-making about the degree of transferability of a given item. This may be one of the reasons why Kellerman's approach was regarded by Meara (1984) as an attempt to establish an "explanatory principle". Gass (1979) has experimentally shown that the learner's acceptability judgements concerning relative clauses reflect Comrie and Keenan's "accessibility hierarchy" (Comrie & Keenan, 1977 and Keenan & Comrie, 1979). Gass suggests that once we have a clear idea of what language transfer consists of, what is or is not transferred, and what constitutes evidence for the presence of transfer, the transfer data can offer a useful means of investigating the role of language universals in L2 acquisition. Gass's experiment shows at least that the learner's judgement about transferability is not random but principled linguistic behaviours. Kellerman's and Gass's experiments relate to Corder's following view:

"...what we must also account for is the relative different speed of learning a different second language in relation to various mother tongues. As has been suggested, the time taken, or the learning task faced, by an infant acquiring any language is assumed to be of approximately equal magnitude. All infants achieve similar levels
of communicative competence at approximately the same stage of development. If then second language learners are engaged upon the task of creating for themselves a grammar of any particular second language, all starting from the same point, and apparently all following the same developmental sequence thereafter, why is it that they typically take different time over the job? Why is it that apparently the same task differs in magnitude for different groups of learners? The hypothesis here proposed states that, other things being equal, (e.g. motivation and access to data, etc.) the mother tongue acts differentially as a facilitating agency. Where the mother tongue is formally similar to the target language the learner will pass more rapidly along the developmental continuum (or some parts of it), than where it differs. Genetically related languages are assumed to share a larger number of rules, particularly in the deep grammar, differing principally in the more superficial aspects. Passage along the developmental continuum is therefore rapid until those relatively superficial distinctions are met, whereas in the case of unrelated (distant) language differences exist along the whole developmental continuum, slowing down the speed of learning. This hypothesis is testable by a comparative study of learners acquiring two different languages simultaneously under the same learning conditions, e.g. exposure, teaching, motivation, etc."


Corder suggests that the transfer data can provide us with research opportunities to look into a relationship between language distance and language acquisition. This issue is now called "psychotopology" (Kellerman, 1983) or the reflection of cross-linguistic generalization upon a learner-language.
The third type of research above is concerned with language transfer, but the way we look at transfer data is distinct from error analysis and contrastive analysis. The new approach does not make much use of objective linguistic analysis. Kellerman's research method illustrates the point (see §4.1). Using multi-dimensional scaling, Kellerman reconstructs the NL structure on the basis of learner-generated data. Then, he tries to discover in what way the learner's judgements correlate with the reconstructed semantic maps. In this sense, his research interest is set in the discovery of learner's judgemental process (i.e., process-oriented research). On the other hand, when we use such linguistic means as contrastive analysis to explain the learner's errors or intuitions, the reader might not be fully convinced of the account we supply, probably because the psychological plausibility for such linguistic analysis is difficult to prove. Contrastive analysis is the investigator's explanation concerning the learner's intuition, and the investigator's explanation does not necessary correspond to a learner's intuition or 'feel' of transferability. We must make sure what constitutes the evidence for the experimenter's putative assumption of the existence of transfer in the learner-generated data. Long and Sato (1984) indicate that "contrastive analysis sought to explain a psycholinguistic phenomenon by exclusively linguistic means..." The linguistic means we adopt are normally 'reference grammars' which are objectively generalized accounts about the well-formed TL forms. We are not yet sure whether these descriptive means is psychologically real from the view point of a learner.

In theory perhaps, learner-language must be investigated entirely in terms of process-oriented and learner-oriented methodology. However, it seems to me that the methodology which aims to be explanatory needs to be combined with linguistic analysis, because reference grammars are part of the formal learner's input. At the same time they are descriptive of the learner's linguistic goal. The IL
researcher may find the two complementary approaches to be useful resources (see below).

Dulay and Burt's "creative construction hypothesis" has been regarded as the antithesis of the older "restructuring" hypothesis (see Corder, 1979). In the light of learner-language, this dichotomy becomes somewhat different (see below). In the extreme forms, the former hypothesis assumes that at the starting point of TL learning, the learner does not have any knowledge to fall back to. The learner "borrows" bits of NL knowledge to make up for his ignorance for the purpose of achieving immanent communicative needs. The latter hypothesis on the other hand assumes that "the starting point for learning was the mother tongue": Corder (1979: 2). These theoretically possible extremes do not occur separately on their own in learner-language. We can therefore assume that these competing hypotheses are complementary to each other. Once we accept a fact that both NL transfer and "recreation processes" (solely based on the TL) occur in actuality, we may even go further and suggest that "recreation process" relates to accommodation and that "restructuring process" relates to assimilation. Corder's definition of accommodation and assimilation is in fact proposed as reconciling the competing hypotheses to accommodate common occurrences of NL transfer and TL-oriented modifications of IL rules:

"In short, the learner, as a result of his interaction with speakers of the target language, is engaged in the task of creating for himself an ever more adequate internal grammar of the language. He does this by the two basic processes of accommodation - adapting his interlanguage grammar to fit the perceived facts of the language, and assimilation - attempting to fit newly perceived facts into the present state of his interlanguage grammar."

Corder, (1978c: 79)

We can notice the conceptual equivalence between Corder's
definition of accommodation on one hand and Dulay and Burt's conception of "creative reconstruction hypothesis" cited earlier (1974). This reinterpretation of recreation and restructuring hypotheses is useful at least in two respects.

(1) Once we regard assimilation and accommodation as occurring both in the acquisitional process of the TL and of transference, the resulting picture of language learning and acquisition we will have may be more realistic.

(2) Once we regard assimilation and accommodation as the basic processes, the other linguistic phenomena, such analogical errors as overgeneralization or simplification seen in pidginization, \( L_1 \) and \( L_2 \) acquisitional processes, might be explained within the same framework. (This may mean that Dulay and Burt's recreation hypothesis is important, although the hypothesis itself has not been experimentally demonstrated very well.) For instance, Anderson's "nativization" and "denativization" is congruent to the present view. Anderson (1979) regards the former as the acquisitional process in which the learner faces his TL in the light of his already acquired NL knowledge and the latter, as the acquisitional process in which the learner has access to linguistic features of the target input.

The above two points (1) & (2) have been in fact suggested by several researchers in the past although the terminologies are different. I will cite or summarize the views representing the position. Selinker (1972) views that a second language learning strategy involves the simplification of the target language system to 'reduce the learning burden': by reducing grammatical redundancy such as third person singular \( s \), the learner makes his task of learning easier. This learning strategy is also used as the communication strategy. The simplification here stands for assimilation. According to Corder's definition, the learner tries to fit complex TL rules into his simpler IL. Taylor (1974) recognizes the connection between these strategies and Ausubel's meaningful learning theory (assimilation theory) (Ausubel, 1963 and 1968) as follows:
"Overgeneralization of target language rules, the omission of redundancies, and a reliance on native language structure when enough of the target language is not known seem, then, to be the most important cognitive processes in second language acquisition.

These proposed cognitive processes or strategies are given a degree of psychological validity when we consider Ausubel's meaningful learning theory (subsumption theory) (Ausubel 1967) and Brown's theory of cognitive pruning (Brown 1972). Both Ausubel and Brown maintain that there is a difference between rotely and meaningfully learned material, and that it is only the latter which is subsequently stored in long-term memory. Meaningful learning involves "subsumption" and "selective forgetting," and is a process by which new material is retained in long-term memory only when it is able to be related to and subsumed by already existing cognitive structures. Meaningful learning helps to explain concept formation: new concepts are meaningfully learned only when they are directly related to and subsumed by previous experience that is already in a "permanent" (i.e., more stable) cognitive structure."

(Author's emphasis added.) Taylor, (1976: 27)

Ausubel's assimilation theory is broad enough to cover the usual notion of accommodation (Ausubel, 1985), since it includes (a) "subordinate learning", (b) "superordinate learning", (c) "combinatorial learning", and (d) "assimilation theory" which Ausubel characterizes as follows:

"New information is linked to relevant, pre-existing aspects of cognitive structure and both the newly acquired information and the pre-existing structure are modified in the process. All of the above forms of learning are examples of assimilation. Most meaningful learning is
essentially the assimilation of new information."  
Ausubel. (1985: 76)

Subordinate learning has two subclasses: "derivative subsumption" and "correlative subsumption". The latter also involves the change and reorganization of existing structure:

\[
\text{Established idea } X \\
\text{New } \rightarrow y \quad u \quad v \quad w
\]

"In correlative subsumption, new information y is linked to idea X, but is an extension, modification, or qualification of X. The criterial attributes of the subsuming concept may be extended or modified with the new correlative subsumption."

Ausubel, (ibid.)

As the above quotations make clear, assimilation does not mean that the language learner transfers his NL in a simplistic and straightforward fashion. It involves modifications and changes. In this sense, Ausubel's notion of assimilation includes our common notion of accommodation. If Taylor's insight into the connection between Selinker's learning strategy and Ausubel's assimilation theory is tenable, Selinker's use of the term 'strategy' above may be rather misleading; as Taylor says, it may stand for cognitive process.

Taylor's view that overgeneralization is the outcome of simplification (and by implication, assimilation) is supported by other researchers.

"Simplification may thus be considered as a universal learning strategy based on the extension or application of rules. Overgeneralization and analogy are instances of the same process."

Richards, (1975: 118)
"The real point, as psycholinguists have reminded us for years, is that previous learning influences subsequent learning and behaviour. This phenomenon can be observed in first language acquisition as Ervin (1964) noted, and it often called overgeneralization. It is perhaps a pity that a special term "interference" was used by habit theorists to describe in second language learning, behaviour that arises from a similar overgeneralization process as is at work in first language acquisition."

Kennedy and Holmes, (1976: 82)

In the context of present discussion, Richards' view suggests that on using a simple IL rule beyond the limit of rule application, the learner has recognized some similarity between a given rule and an over-extended item. In this sense, assimilation involves the learner's intuition of similarity. This may be the reason why Richards regards over-generalization and analogy as instances of the same process. According to Ausubel, "meaningful learning takes place if the learning task can be related in nonarbitrary, substantive .... fashion to what the learner already knows" (1968: 24). The linguistic input to a learner will be turned into "intake" only when he is ready to internalize it (Corder, 1978c). In Krashen's terminology, acquisition takes place when the "comprehensible" input is available to a learner (see §1.1). This cognitive readiness on the part of a learner was recently found to be required even in the area of NL transfer. The learners do not transfer their NL knowledge until his IL development attains certain conditions. Wode writes as follows:

"But the specific pl. formation rules of German were not carried over to English. However, quite a few of the German regularities relating to the positioning of the negatives were apparently carried over to English. This seems to imply that certain conditions have to be met for what is commonly called interference to take place
at all. Therefore the notion of interference has to be developmentalized if it is to provide any fruitful insight."

Wode, (1976: 27)

This suggests that the learner does not regard an NL item as transferable unless his IL competence reaches a certain stage. He will, then, regard the NL item as assimilatable to the TL system. Wode (1978) calls this "crucial similarity measure" (1978: 116). Krashen (1983) supports Wode's view as an important factor to explain language transfer in language learning. Zobl (1980a and 1980b) examined this "selectivity of L1 transfer", experimentally and obtained the confirmation of his hypothesis.

If accommodation and assimilation represent two basic learning or acquisition processes, investigations into language transfer would offer us experimental opportunities to study the processes of learning and acquisition processes. In this connections, there are three interesting studies, two in phonological developments (Dickerson (1975) and Dickerson (1976)) and one in lexical development (Strick, (1982)).

Dickerson (1975) and Dickerson (1976) demonstrated that the acquisition of L2 phonology was largely a restructuring process. The studies showed that the Japanese subjects started with replicating Japanese phonological system in pronouncing English words and gradually developed a TL-like phonological system. Interestingly, they found that "variable rules" used by Labov in his synchronic and diachronic studies of sound changes could describe the L2 learner's variable performance adequately. Dickerson and Dickerson modelled the learner's developmental patterns on Bailey's "wave mechanism" as well. The discussion shows that Labov's finding concerning the synchronic reflection of diachronic changes is obtained by the L2 data involving the phonological acquisition. (Present study also gives some examples of synchronic reflection of diachronic change of meanings (see chap.3 and 10))

Strick (1982) found a restructuring process in lexical
development. Using a multi-dimensional scaling technique, three semantic maps were reproduced: one obtained from native speakers of English and two from Iranian native speakers and Iranian adult learners of English. He compared their use of address terms. And Iranian child learners of English also participated in this experiment. The semantic maps produced from the two native speakers were used as reference maps. The semantic map among the Iranian native speakers had two dimensions of 'sex' and 'status' and the semantic map among the Americans, two dimensions of 'sex' and 'intimacy'. The use of address terms among adult learners of English showed the greater amount of reliance of 'sex' dimension than 'status' dimension. The same tendency was observed among the child learners of English. These findings were interpreted as indicating a progression from the "perceptually-based" dimension to the "abstract culturally relative dimension" in the second language.

In the area of vocabulary learning, Blum and Levenston have investigated lexical simplification. I will refer to their experiment to discuss one difficulty in a lexical study concerning simplification or assimilation. In their experiment (1980), they used a discourse completion task among native speakers of English and L2 learners. This experimental design is motivated by the following assumption.

"Besides the need to memorize vocabulary in the L2, the language learner is faced with the immense task of internalizing the exact nature of these inter-relationships between lexical items (Richards 1976: 81). The learner's need to simplify is thus explained by the complexity of the task of acquiring command of all aspects of the native speaker's competence; his ability to simplify derives from his own semantic competence in his first language."

Blum and Levenston, (1980)

They tried to find out whether native speaker's use of
simplification is similar to L₂ learner's simplification. Their notion of simplification covers the use of avoidance, superordinate terms, approximation, synonymy, transfer, circumlocution and paraphrases: Blum and Levenson, (1980a:309). In Blum & Levenson (1980) and Blum-Kulka & Levenson (1983), in reference to simplification used in L₁ speech and published texts, they exemplify the point that L₁ speakers also use simplification when talking to a foreigner, addressing other native speakers and in writing. They intend to show that the IL notion of transitional competence is applicable to the sociolinguistic level in the TL society. (Corder recognized parallels between IL speech and in the native speaker's use of deviant forms (see chap.1). More recently, Tarone (1982) developed a model which incorporates Blum & Levenson's research aim of style-shifting along an interlanguage continuum of styles.) Blum and Levenson argue that "strategies of simplification are basically universal in nature" (1980b:405), but according to Meara (1984) their experiment does not appear to demonstrate their experimental aim. Meara points out the difficulty of research dealing with vocabulary learning within the framework of interlanguage:

"...there is no way that lexis can be sensibly handled using the tools and methods which have been traditionally associated with interlanguage studies, and for interlanguage to attempt to study lexis would result in turning it into something else. This dilemma is, of course, more apparent than real, however. It rests on an underlying assumption that the current tools and methods of interlanguage form a definitive canon, and not just a collection of utilities. At bottom, the problem is a simple confusion between aims and methods."

Meara, (1984: 230)

Meara's view derives from the fact that words have been fully researched in psychology since the end of the last century. Any IL researcher cannot ignore such well-
established findings in psychology. Meara advises to a lexical IL researcher to pay attention to the psychological findings and to contrive a well-controlled experiment. If we are interested in cognitive mechanisms, experiments in cognitive psychology must be consulted so that our experimental method is in line with our research aim of cognitive mechanisms. However according to Smith, the heavy reliance on the previous psycholinguistic studies must be cautioned:

"Hence the budding lexical interlinguist would be making a big mistake by turning exclusively to the psycholinguistic literature and ignoring the linguistic work on matters lexical..."

I would argue that a move to make research more fundamental by adopting the methods and techniques of experimental psycholinguistics will only be worthwhile if accompanied by a serious consideration of the theoretical underpinnings.


It looks as if the combination of Meara's view with Simth's advice offers a safe guideline. I have regarded the two approaches as essentially complementary (see above). But, if we are interested in finding out what gives rise to the fact that certain items are acquired earlier than the others, we must pay proper attention to what psycholinguistics have established previously. The performance analysis allows us to talk about the order or the rate of acquisition but it does not enable us to conclude anything beyond the order or the rate of acquisition.
Chapter 3: an operational definitions of word-meaning

§ 3.0 introduction

This chapter aims to present an operational definition of word-meaning and an operational distinction of competence vs performance in word-acquisition in preparation for the analysis in later chapters. The definition and distinction of these terms will be useful in explaining the analogical process which this research puts forth occurring in the judgemental process of translatability. (The relevance of these precise definitions and distinctions is limited to this research, but, are vital in explaining some of the findings obtained. The discussion in this chapter will demonstrate how this experimental study dealing with only two words can illustrate a general issue in L2 vocabulary learning."

In order to meet the guidelines mentioned in §1.1, the following pre-cautions were taken. (a) The previous experimental studies on which we base our operational definition have been tested both among L1 children and adults. (b) The previous studies have used both artificial and natural materials (ie., linguistic) as the experimental items.

The analogical process includes "over-extension" among L1 children. When a child recognizes some similarity between two objects to which an adult would apply different lexes, the child uses the same word for the two objects, if he has not acquired the word for the other object yet. This recognition of similarity between two objects on the part of the child is analogous to the L2 learner who only knows one TL word for two NL items and over-generalizes the TL word to make up for his ignorance. Just like the child in our example, because the learner recognizes some similarity between the two NL items, he ventures to over-generalise the use of the TL word. In this sense, there is some similarity between over-extension among L1 children and over-generalization among L2 learners.
For this reason, I take up three well-known lexical hypotheses in \( L_1 \) child language studies and see which of these explains over-extension best. The best hypothesis would be more applicable to the \( L_2 \) situation. The discussion in §3.1 supports Rosch's prototype hypothesis (PH). Since this conclusion has been accepted by specialists in \( L_1 \) studies, the nature of the present discussion is a summary of established views (see §3.1). It is noteworthy that PH has been examined whether it can account for lexical acquisition not only among adults but also among \( L_1 \) children (Rosch et. al (1976) and Mervis & Pani (1980)).

Lexical acquisition among children can best be likened to experiments on concept-formation in psychology. The relationship between \( L_1 \) lexical acquisition and concept-formation is also analogous to the relationship between \( L_2 \) acquisition and the formation of hypothesis (see §1.0). The function of concept-formation in \( L_1 \) acquisition may be illustrated by the following familiar anecdote.

When a child correctly recognizes a dog he has never seen before as a dog and calls it 'doggie', Bruner (1957) argues that the child has manifested an ability to generalize from his previous experience of seeing many dogs. Since the perceptual patterns which the category 'dogs' has differ in their physical attributes, the child has learned something that enables him to classify a particular instance he just saw into the category 'dogs' and to connect his 'dog-concept' with a verbal sign 'doggie'. This something which the child acquired is called 'concept' in psychology, and psychological experiments in this domain are known as the study of concept-formation or the formation of hypothesis. It is important to bear in mind that the presence of 'dog-concept' in cognition enabled the child to identify the animal as a dog and his one-word utterance is evidence of the existence of the internal concept.

The above example shows the close connection between perception and conception. Bruner suggests that there
is a "continuity" in the rules of inference used at both perceptual and conceptual levels. According to Bruner (1957), the acquisition of concept involves four factors. The child (in the above example) needs to know (a) the "critical" attributes which visual objects belonging to the same category have, (b) the manner in which these attributes are combined, (c) the "weights" assigned to various attributes which enable the child (subject) to recognize and identify a specific instance as the members of a category and (d) the "acceptance limits" of each category. These four points have been experimentally studied. The more recent experiments have confirmed the relevance of Bruner's four factors which are captured by the PH.

The above use of the term 'concept' could be associated with "intension" in semantics. But Lyons (1977:208) indicates that the 'concept' used in child language studies is more properly regarded as "denotation" rather than "intension". The main reason behind Lyons's observation appears to be that we merely infer 'intension' on the basis of 'referents' which the child applies the same word to; but, this 'intension' inferred by the outside observer may not correspond to "intension" in the mind of the child. More important, logicians' use of the term "intension" is a "dual" concept of another theoretical entity of "extension". From the theoretical view point of semantics, 'referents' belong to "denotation". In this way, the denotational meaning is distinct from the intensional meaning. We need to bear in mind that the discussion in this chapter relates to the denotational meaning and that for the same reason, the operational definition which will be offered (in §3.4) may be confined to the denotational meaning of words.

§3.2 outlines the view that PH is an important development of the earlier works by Bruner, Posner, Keele, Reed, etc. Posner and Keele's experiments (see below) are particularly relevant in lexical acquisition. There are several empirical studies which indicate that the child can acquire
a concept \textit{instantaneously} ( see below ). These reports fit in with the findings obtained by Posner & Keele (1969, etc.) and Reed (1973). In their experiments, they used artificial materials such as randomly arranged dot-patterns or schematic faces. Because of these experimental materials, their experiments vividly demonstrated the surprising human capacity of abstraction; the subjects were capable of abstracting the prototype (the central tendency) on the basis of the 'poor' examples they saw during the experiments, without having ever seen the prototype itself.

Rosch’s PH is notionally stated and ambiguous. For instance, she generalizes her experimental findings into the two key concepts: a "basic level" which is defined as the "most inclusive member", and "family resemblance" which means that the members of a category are all connected in terms of similarity. Mervis and Pani (1980) offers some clarification to these concepts in terms of "maximal within-category similarity" and "minimal between-category similarity". But these definitions become only meaningful by looking at the previous experiments upon which Rosch and her colleagues based their experiments. By connecting their definitions with the earlier works, the ambiguity in PH is clarified. PH involves not only the abstraction of prototype but also the abstraction of acceptance limits or judgemental limits. In other words, the subjects abstract a prototype as well as a membership (specific members belonging to a prototype or the range of lower and upper acceptance limits).

§3.3 refers to the experiments by Rumelhart and Abrahamson (1973) and Sternberg (1977). The purpose is to see whether the present understanding of PH accords with their experiments on analogical reasoning.

§3.4 begins with the classical analogical paradigm which is adopted by Rumelhart and Abrahamson. This paradigm has been utilized by historical linguists to explain historical changes of meaning. On the basis of this
paradigm, I summarise four types of L₁ over-extension treating them as prime cases of lexical performance in relation to Stern's (1931) or Waldron's (1979) diachronic perspective of changes of meanings. The discussion suggests that both analogical reasoning and chained associations occur diachronically and as well as in a short span at a very early stage of one's life. In this sense, there are parallels in diachronic changes and these brief changes at a personal level.

In the course of discussion, we arrive at the simple criterion of all vs some relevant features for the purpose of the present distinction between competence and performance. I use the framework of 'distributive lattices' derived from Piaget's INRC operations which are suitable for some discussion of assimilation and accommodation in chap.10. We show that Ullmann's characterization of word-meaning accords with the present operational definition, showing at the same time that the notional definitions of "maximal within-category similarity" and "minimal between-category similarity" can be accommodated by the present operational definition of word-meaning.

§ 3.1 three hypotheses in L₁ acquisition

This section takes up Clark's proposal of Semantic Feature Hypothesis (SFH), Nelson's Functional Core Hypothesis (FCH) and Rosch's Protype Hypothesis (PH). Additionally, Barrett's contrastive hypothesis (CH) is briefly touched on.

There are three basic assumptions in SFH (Clark, 1973).

(a) The meaning of a word can be decomposed into semantic features. These features form a hierarchical dependency structure.

(b) The initial semantic features the child acquires are partial, as opposed to full adult meaning. They are mainly perceptual.

(c) The process of acquiring these semantic features is gradual. The child gradually increases the semantic
features until they correspond with those in the adult's words. Starting with a general semantic feature, the child acquires particular features which are relevant to individual words.

Let us take an example Clark gives of the child who uses the term, bow-wow, referring to dogs. Immediately or after some time this bow-wow is overextended to refer to dogs, cats, horses, cows, and sheep. (Some children requires a silent period; according to Bowerman, it takes a couple of days or more than one month before they actually start to use the words they have heard before: (1978:251)) Since 'bow-wow' covers five kinds of animals, Clark infers that the semantic feature which the child is likely to have acquired is highly general in meaning. When the word, moo, is acquired in reference to cows, other features are added, so as to keep the meaning of moo separate from bow-wow. Thus, as the child acquires a word for each member of the 'bow-wow' set, some specific features that differentiate the meaning of one word from the others is added to the child's repertoire of semantic features. For this reason Clark regards the development of semantic knowledge as from general to particular.

Clark's three basic assumptions predict that the acquisition of positive terms will occur earlier than that of negative terms. Several experiments demonstrated that when there is a pair of words describing a spatial relation, one of the pair is perceptually preferred, e.g. up is preferred to down (cf. H. Clark & Brownell, 1975, Olson & Laxar, 1973, Rudel & Teuber, 1963, H. Clark & Chase, 1972, 1974 etc.) Since perceptually preferred terms are defined as positive or unmarked words, they relate to a natural predisposition (cf. H. Clark, 1973, Reudel & Teuber, 1963). Therefore, assumptions (b) & (c) naturally lead to the prediction that positive terms are acquired earlier than negative terms. This prediction has been confirmed by a variety of elicited data (Donaldson & Balfour, 1968, Donaldson & Wales, 1970, E. Clark, 1971, C. Chomsky, 1969, Piaget, 1928 etc.) In this sense, SFH is supported by
these experiments.

Further, on the basis of 27 diary studies of diverse languages, Clark generalized the following point:

1. Over-extension is language independent and occurs universally within the same age-range (1;1 & 2;6 years).
2. The child's frequent question, What's that? marks the end of the early form of single word over-extension.
3. The features such as movement shape, size, sound, taste and texture are used criterially in overextending the use of words. The majority of overextensions appear to be based on perceived similarity.

According to Greenberg & Kuczaj (1982), Clark's basic assumption (a) presupposes an analytical ability on the part of infant. There are some experiments which demonstrated that the analysis of criterial feature abstraction is a complex advanced ability which should follow rather than precede holistic understanding (Posner, 1973, & Hyman & Frost, 1974). In this connection, Garner made a useful distinction between "integral" and "separable" stimuli in order to reveal the interaction between input structure and the forms of processing (1970 & 1974 see also chap.4 & 9). In support of his argument, the experiments involving adult subjects demonstrated that the differences in the structure of input stimuli do lead to different forms of processing (Garner 1974, Garner & Felfoldy, 1970, & also chap. 4 & 9). "Integral stimuli", those such as colour whose dimensions of hue, brightness and saturation, cannot be perceived as "separable" under normal circumstances, are processed holistically in terms of the 'feel' of similarity and do not permit selective attention. Whereas, separable stimuli such as the various sizes or angles of a radial line are perceived as perceptually distinct components and they are processed in terms of each dimension; therefore selective attention is possible. These different forms of processing do occur young children and appear to have major developmental implications. The experiments by Shepp & Swartz (1976), & Shepp (1978) show that, presented with certain combinations of attributes,
young children appear to treat practically all attribute combinations as if they were integral. Smith & Kempler (1977 & 1978) also supported this tendency. In their experiment 5 year-olds (the youngest children tested) preferred to classify items by means of overall similarity rather than on an analytical basis of the presence of the same value on a particular dimension. These experiments indicate that integral processing based on global similarity is a preferred form of processing among young children.

Nelson argues against Clark's conclusion (3) and, to some extent, also her basic assumption (b), that the core of the child's concept of a word is functional. According to her Functional Core Hypothesis (FCH), perceptual similarity is employed merely to identify objects and that functional similarity (rather than perceptual similarity) is criterial for the over-extension of words (1974, & 1977). Nelson regards the functional properties as defining intension and the perceptual properties extension. (As we have seen earlier, from overt behaviour one can only infer what is present in cognition. We may only grasp putative 'intensional' properties, but they may not correspond to "intension" in semantics. One can avoid this danger by adopting the term, 'denotation' as Lyons suggests.) However, Nelson's operational definition of function in her experiments (1973) involves perceptual aspects such as the change of state from wet to dry, various shapes and colours. The experiments by Bloom & Lahey (1978) and Rosch et. al (1976) show that function and perception are correlated and fundamentally indistinguishable. Likewise, Bowerman (1977 & 1978) presented empirical data to show that, while perceptual similarity appeared to be the major basis of overextension, overextension based on function was minimal, but not nonexistent. As Bowerman indicates, identifying objects on the basis of perceptual similarity is also part of the process of acquiring a word.

Clark's conclusion (3) has some difficulty in explaining fuzzy word boundary porposed by Lakoff (1972). Palermo (1982:338), and Greenberg & Kuczaj (1982:228) indicate
that according to SFH, a criterial feature defines the range of referents of a word; as long as referents have the criterial feature, all of them are equal in terms of their membership of a category. In this sense according to SFH, a word has a definite word boundary. But this view is contrary to Lakoff's linguistic analysis as well as the following experimental studies. Labov (1973), Andersen (1975), Rosch (1975), Fernandez (1972), Oden (1975), and McCloskey & Glucksberg (1979) show that word-boundaries are fuzzy and not well defined. These studies demonstrate that while the subjects show substantive agreement as to whether one instance is the 'best exemplar', they judge other instances as having varying degrees of 'exemplariness'.

SFH predicts over-extension more readily than underextension, since a general feature is assumed to be acquired earlier than a specific one. However, there are at least some studies which contradict this prediction. Gruendel (1977) found a developmental progression that a word is used for a specific referent at first, and then used as a symbol for a general class of objects. Reich's anecdotal report gives one such example (1976): a child used the word 'shoes' only in reference to his father's shoes in the bedroom before he began to use the word for any pair of shoes. Anglin (1977) has found that some words are initially underextended (supporting Gruendel's finding above) but that other words are from the beginning overextended. Since in these longitudinal studies only a small number of children (one to three children) were dealt with, the problem of individual differences may come into this issue. But, the fact that SFH cannot explain underextension adequately makes SFH rather unsatisfactory as a model.

Thus, SFH is not satisfactory for three reasons;

(a) It relies heavily on the analytical ability on the part of infants.

(b) It cannot incorporate adequately the idea of fuzzy word-boundaries nor the degree of category membership.
It does not account for underextension which is an important phenomenon often documented in empirical studies.

Barrett's proposal of "contrastive hypothesis" (CH) explains the process in which a child corrects the use of over-extended or under-extended words to match the adult's use of the same words. Barrett argues: "the meaning of a referential word cannot be acquired by the abstraction from the referents of a word, but by an abstraction of the contrasts between positive and negative referential instances" (1978:207). According to Barrett, CH is necessary for the child to acquire the range of application of referential words. Winston (1973) also indicates from the viewpoint of artificial intelligence the necessity of "negative information" as a means of correcting erroneously learnt knowledge. CH in a sense relates to this notion of "negative information". The word which the child contrasts with another word in order to correct the use of the former serves as "negative information". The contrastive word negates some aspects of the child's erroneous usage. This negative information as a source of CH does not have to be supplied to a child deliberately. As Chomsky (1981) indicates, when the adult does not use a certain word in cases where the child expects it to be used, these contexts function as a kind of negative evidence for the child. This naturally available negative information may be sufficient for the child to correct his over-extension and under-extension. As we have seen above, both SFH and FCH contain the essential point of CH, although they do not emphasize it. In this sense, CH is relevant for the purpose of this chapter.

Rosch's Prototype Hypothesis (PH) has been regarded as promising in research on L1 lexical acquisition (e.g., Mervis & Pani, 1980). PH can be characterized as follows.

(a) There is "a basic level" for a group of words which are subsumed under one superordinate noun (i.e., category name): e.g., apples, oranges, melons and so on are the member of the category 'fruit' (see Rosch,
1973 & 1975b)

(b) These items belonging to one category have "family resemblance". (see Rosch & Mervis, 1975)

(C) The basic level which is sometimes called 'the best exemplar' or "prototype" is characterized as "the most inclusive member". It is defined by the two concepts of "maximal within-category similarity" and "minimal between-category similarity" (see Rosch, 1978 & Mervis & Pani, 1980)

Greenberg and Kuczaj (1982) point out that "the occurrence of under-extension and over-extension is a function of similarity" (1982:306). Greenberg and Kuczaj's use of the term similarity includes not only perceptual but also functional similarity. This suggests that PH is in keeping with both SFH (which stresses perceptual similarity) and FCH (which stresses functional similarity). PH is also more suitable in explaining over-extensions and under-extensions which are based upon perceptual and functional similarities. Furthermore, PH claims that there is a family resemblance in semantically related words. This position accepts a "fuzzy" word-boundary more comfortably than SFH and FCH do. Thus, PH is more general than SFH and FCH, and it can account for what SFH and FCH had difficulty in explaining.

As for the "basic level", the notion was proposed by Brown as early as 1958. According to Clark & Clark (1977), the "basic level" relates to Berlin's "generic level" in anthropological linguistics (Berlin, 1972 & Berlin et. al 1968). Bruner et.al also talked about the "typical instance" (1956: 64). Posner & Keele (1968) and Attneave (1957) demonstrated that the subjects could abstract prototypes without ever seeing prototypes themselves, suggesting that prototypes might have some mental reality. In this sense, Rosch and her colleagues substantiated the seminal idea of "basic level" by running many experiments.

Rosch and her colleagues' experiments are more directly related to linguistic materials than the previous experimen-
tal works (see § 3.2). Rosch (1973 & 1975b) asked the subjects to rate words belonging to several superordinate categories on a 7-point scale in response to the question 'how well each word exemplifies a given superordinate category. The results showed a significant agreement among the subjects with 'murder' overwhelmingly rated the best instance of 'crime', and 'apple' almost always the best fruit and so on. In another experiment (1975c) Rosch demonstrated that the rank order of typicality ratings could predict reaction times for the probe questions such as "A car is a vehicle" vs "A tank is a vehicle". It showed also that the questions with best exemplars were verified more rapidly than those with peripheral (as opposed to central) members. In Rosch (1975a), the subjects were given a superordinate category name as a prime and then asked to judge whether a pair of words or pictures were the same or not. For example, the word sport was given just before the pair, football/football. This is a priming technique originally used for a letter-matching paradigm in which subjects are supposed to respond as quickly as possible whether a pair of letters visually presented are same or different, "same" being defined as physical identity (e.g., (AA)) or as having the same name (e.g., (Aa)). It is known that response time improves when some relevant information is provided beforehand, eg a prime with letter 'A' prior to the presentation of (AA) or (Aa). It is known that the subjects remember uncousciously not only how the letter used as a prime was written (e.g., capital or small letters) but also how the experimental stimuli were coded in making responses (Beller, 1971 and Posner, 1969). Based on this finding, Rosch hypothesized that the coding process would distinguish whether the subjects conceived the experimental items as holistic prototypes or as a set of analysable features. She assumed that there would be two kinds of coding processes: the similarity-based global coding and the analytical checking of features. She predicted that the former coding would be more suitable for prototypes and near-prototypes. The same responses were expected to be easier when a pair had more features in common. These predictions were supported by the data.
and the same effect was observed in both word-matches and picture-matches. The effect was most consistent for the best exemplar. On the basis of these findings, Rosch argued that on hearing a category name, we would evoke the prototype of that category rather than a list of constituent features. She concluded that the "cognitive representation of a category" was likely to be a "concrete image of a central category member" (1975a:25). (Incidentally, we can note that Rosch's idea of prototype appears to be associated with a holistic grasp rather than a collection of features which can be analysed if one wishes.) In subsequent experiments (Rosch & Mervis, 1975) they attempted to identify at which level of the hierarchy does the best exemplar with a concrete image exists. They asked the subjects to list as many attributes and action sequences as they could which were associated with the words at each level of the hierarchy. They found that the words at the basic level between super-ordinate and subordinate levels have the most common attributes and action sequences with other members, and the fewest attributes in common with other categories. This finding is known as "maximal within-category similarity" and "minimal between-category similarity". Rosch et. al (1976) replicated these results among a large number of subjects and also found that "basic level objects" were the most inclusive level "at which the experimental objects depicted in outline were identified." At this basic level a "large increase in objective similarities in shape" was recognized.

We have outlined the three features of PH so far. All these features of PH do not explain how a prototype (or, the basic level) is formed. Rather, PH is concerned with what the structure of word-meaning is in psychological terms. For this reason we adopt PH as our operational definition of word-meaning and PH explains over-extension which is characterizable as a "function of similarity" (see above). In this sense, PH is suitable for investigating analogical process we are investigating. As I have said previously, PH is an important development which captures the main psychological findings by Bruner,
Posner, Keele, Reed and so on. Their experiments directly relate to the formation process of prototype. Since present research is concerned with this formation process, we will examine this below (§ 3.3). Before we proceed to the next section, we will outline two views about the basic level.

1. The basic level itself is language-independent.
2. The outcome of prototype formation depends on the context in which it develops.

Brown (1976) states that the basic level has a developmental implication. Brown argues that what is basic to a child may be superordinate for an adult; and this difference is studied by Mervis & Mervis in relation to Mother's talk (1982). Dougherty (1978) points out that the basic level may differ across cultures. Observations of these kinds suggest that the basic level is context-dependent and that the actual content of basic levels depends upon the level of cognitive maturation. This suggests further that the basic level is an abstract functional generalization about the outcome of prototype formation process. It may follow that, as Rosch (1978) argues, the basic level is "abstract foci" in which prototypes are formed with more ease than at any other levels. If so, PH should show that there is a significant effect of the basic level upon learning. This assumption was in fact experimented by Rosch & Mervis (1976) and Mervis & Pani (1980). The experiment in chap.10 (see also appendix2) also demonstrates that words at the basic level are learned in a more stable way.

Rosch and Mervis (1976) and Mervis and Pani (1980) used artificial categories among adult subjects. Based upon the premise that people generalize on the basis of similarity, they examined the effect of prototypes (the basic level) upon learning. They found that an initial exposure to prototypes (i.e., the exemplars with maximal within-category similarity) facilitated learning in terms of speed and accuracy. Whereas exposure to poor exemplars did not promote learning. They also found that the subjects learned prototypes as members of a category long before
the poor exemplars. Further, when Rosch et al (1976) re-analysed Brown's well-known longitudinal data on L1 acquisition, they found that the object names the subjects used at Stage 1 were essentially the words at the basic level (1976:424-5). Moreover, Mervis and Crisafi's experiment (1982) indicated that young children acquired categorization skills in the order of basic, superordinate and subordinate. What emerges from these results does indeed suggest that the underlying process does not differ in spite of the difference of age, materials, not to mention the difference of languages.

§ 3.2 previous experimental works on PH

This brief review is relevant here for two reasons.

(1) I am interested in finding out whether learning occurs instantaneously or gradually. Several empirical studies indicate that lexical acquisition can occur instantaneously (for the "instantaneous model of language acquisition", see Chomsky, 1975). This instantaneous acquisition is explained by the previous experimental studies which Rosch substantiated as PH.

(2) I am interested in the formation of hypothesis among L2 leaners. This means in the present context that we are interested in the process of prototype formation. PH describes the structure of word-meaning in notional terms, but PH itself does not clarify how a prototype is formed in cognition. The following experiments demonstrate very vividly that the subjects abstract prototypes. First, we will look at some empirical studies.

There are several studies which support the "instantaneous" lexical acquisition (Leonard, 1976, Nelson & Bonvillian, 1978, Oviatt, 1979, Ross et al, 1980, and Bowerman, 1978). The first four studies indicate that after a single brief training of one exemplar, infants made progress in learning a concept and its label, applying the label to novel instances. These findings show that children can form a concept instantaneously on the basis of a single positive
exemplar. This suggests that CH mentioned above may not be needed in all cases. It also suggests that instructors or teachers do not need to expose children to a number of different exemplars for children to acquire a concept. The above experimental studies do not say that they dealt with exceptionally talented children. But this surprising human ability of abstraction exhibited merits some attention on our part.

It must be noted incidentally that the instantaneous assumption depends upon how we interpret the structure of the prototype itself. As we see below, there are two hypotheses within PH that are contrary to the instantaneous assumption: (a) "frequency hypothesis" and (b) "averaging hypothesis". Both hypotheses require the subjects to be exposed to many exemplars or one exemplar many times until they can average features or they can abstract a concept as a result of frequent exposures. The averaging hypothesis is for the sake of mathematical convenience on the part of experimenters at times. In the case of figural patterns the central tendency can be equated with an algebraic average. For this reason, researchers often included an averaging model as representing a kind of central tendency (i.e., prototype) (cf. Reed, 1972). In connection with this, there are some interesting studies. Cohen and his co-workers demonstrated that 30-week-old infants form a concept of human face based on several varying exemplars, (Cohen & Strauss, 1979 & Strauss, 1979.) Strauss interpreted this as indicating that an infant uses an averaging procedure to form a prototype for the category 'human face'. However, Schwartz and Day (1979) have shown that 8-week-old infants have the sense of shape constancy, that is, the different orientations of the same object are categorized by the infants as the same object. If they were averaging the various retinal images of a shape viewed from different angles, it would be impossible for them to establish a sense of shape constancy. (If the newly born babies adopted the averaging procedure, their concept of 'human face' would be a human face tilted at the average angle.) The experimenter reported that the infants showed
evidence of knowing the relations among the shapes rather than averaging all the exemplars presented. This implies that averaging or frequency hypothesis is less adequate than it appears.

We will briefly examine the classical studies related to PH. In Posner & Keele's experiments (1969), during training the subjects were presented with visual patterns distorted from such base patterns as a triangle, the capital letters of M & F and a random figure. Each pattern was composed of 9 dots and the degree of variation of each dot determined the amount of distortion from the prototype (the base pattern) which was manipulated by such statistical methods as "random walk" and so on. The set of experimental items was so distorted that we would have great difficulty in recognizing any prototype. One remarkable feature of the results was that the prototypes which their subjects had never seen during the original training were classified into the correct categories just as readily as the originally learned distortions and that the prototypes were recognized by the subjects as the best exemplars. Posner and Keele interpreted these results as indicating that the subjects were capable of abstracting a central tendency which they had not seen before, and forming a prototype during the training sessions. This accords with the empirical finding among L1 children mentioned above. Just as the children managed to acquire a concept after a single exposure, the adults developed the concept of a prototype which they had not seen before. These two findings are similar in that both show the remarkable capacity of human abstraction.

Posner and Keele regarded the formation process as holistic, since their method of distortion was global. This observation reminds us of Rosch's characterization of prototype which she regarded as global in nature. This global understanding corresponds to Garner's "integral" processing, as opposed to "separable" processing (see chap.4 & 10). One might argue that, in the case of Posner and Keele at least, the nature of experimental items might have forced
the subjects to adopt integral rather than separable processing. If the experimental stimuli in Posner & Keele's experiment had been different, the subjects might have used separable processing. We might wonder whether the formation of prototype might never have occurred if the experimental stimuli had been different. This is an arguable point. But there is an experiment by Frank & Bransford (1971) which indicates that it is not the case. Frank and Bransford deliberately used structured prototypes composed of a square, diamond, heart and triangle. These discrete, easily distinguishable visual attributes were combined to create a prototype. Further, without using a method of randomization as a distortion rule, Frank and Bransford employed the obvious distortion rules that were discernible to the subjects. In spite of these deliberate preparations, their results confirmed Posner & Keele's conclusion. Although the training set consisted entirely of distorted patterns, Frank and Bransford discovered that prototypes received the highest positive recognition ratings and that these ratings were in the order of increasing transformational distances.

In Posner and Keele's experiments, two of the base patterns were S and F. Except for the base of random dot pattern, S and F were familiar to the subjects. Although the base patterns were never presented to the subjects and these familiar letters were highly distorted, they were still probably somewhat familiar to the subjects. Peterson et. al (1973) examined the effect of familiarity upon the formation of prototype. They used both meaningful (familiar) and meaningless (unfamiliar) patterns. The results showed that a factor of meaningfulness did not yield any reliable effect (1973:383-4). So far, we have seen that Posner & Keele's and Keele's experiment illuminated PH excellently.

Frank and Bransford's experiment, however, was questioned by Neumann(1974). Neumann demonstrated that when experimental patterns have clearly distinguishable components, the degree of familiarity (i.e., how often the subject is ex-
posed to a particular component or the combination of components) affected performance. But, Neumann's experimental result does not necessarily contradict Posner and Keele's experiment. If we are repeatedly exposed to any item which we are inherently capable of abstracting the main features at first glance, we will retain its memory better than in the case of a single exposure. But whether the item is repeatedly presented or presented only once should not make much difference to the fact that we know the concept. As we have seen already, children can indeed form a concept on the basis of a single exemplar. The single exemplar enabled them to generalize the learnt concept to similar objects of the kind.

Neumann's reanalysis of Frank & Bransford's data has one interesting consequence. According to Neumann's reanalysis, the prototypes used in Frank & Bransford's experiment have the greatest number of attributes and attribute combinations in common with the other members belonging to the same category as the given prototype. This indicates that the prototypes which Frank and Bransford used fits in with Rosch's definition of "basic level" and "best exemplar" (see her basic assumption (c) above). In fact, one of her experiments (Rosch, 1975) adopts this combination of attributes and discrete attributes. When the prototype is defined as the most inclusive member which has the most attributes in common with the other members of a category, it should be remembered that the prototype naturally possesses maximal similarity to the other members of a category.

Both Neumann's and Frank & Bransford's experiments seem to imply another important point. That is, a prototype does not necessarily lend itself to global representation. The properties of a prototype may possess some explicit combinatorial structural relations. This may be particularly so with an adult subject who can attend to attributes separately. This line of discussion returns our attention to the analytical proposal offered by E. Clark and Bruner reviewed earlier. In fact, PH and SFH are seemingly com-
peting theories but they provide us with complementary descriptions of the same acquisitional process of concept. Reed's experiments (1972) illustrate this point precisely, because his 'weighted features prototype model' is syncretic in one important sense.

In the series of experiments Reed used schematic faces. The positions of an eye and a nose, the length of a nose and the height of a forehead were changed to create an experimental set. Thus, the material itself is both holistic and structurally relational. Since similarity measure is convertible into distance measures (cf. Shepard, 1962, Kruskal, 1964, etc.), distorted patterns were regarded as differing from the prototypes according to distances along a similarity dimension. This way of computing the differences of various exemplars enabled Reed to estimate the gross characteristics of the set of exemplars as well as the degree of variability of a category. The results suggested that while the subjects formed a prototype to represent a category, and thereby to classify test-patterns on the basis of their similarity, the subject's sense of similarity was determined by criterial features which discriminate categories. This shows that the subjects abstracted the central tendency and estimated the variability of a category, and identified the criterial features at the same time. Since Reed's "weighted features prototype model" incorporates the assumption of criterial features seen in Clark's SFH as well as the three characteristics of Rosch's PH, we may say that the weighted features prototypes model combines these two theories into one.

Furthermore, the "weighted features" may represent perceptually salient features which Clark regarded as the initial semantic features the child acquires and as the main source of overextension. In Reed's model, the global understanding of similarity (which conveys Posner and Rosch's idea of prototype) is combined with perceptually distinct attributes; (the "weight features" presume that attributes are first of all distinct and permit of separate attention).
Some fore-grounded distinct features in the background of global similarity or one salient feature consisting of a bundle of integrated features (see eq. 6 in chap. 9) is certainly compatible with the usual notion of salient criterial features. In this sense, Reed's interpretation of prototype is not only comparable to SFH, but also, it suggests that the structural representation by means of combinatorial properties is compatible with the notion of prototype (see §3.4 blow).

In the course of this discussion we have seen that the abstraction process involves more than acquiring a prototype. In Posner & Keele's experiment examined above, they also tried to find out whether the abstraction process is centred on the formation of a prototype and on nothing else. They hypothesized that if during the training session, the subjects were only abstracting a central tendency, training in low variability would facilitate the formation of a prototype, while training in high variability would make the process of prototype formation more difficult. Their results failed to support their prediction. The subjects were better at categorizing highly distorted patterns when they had been trained on moderately distorted patterns than when they had been trained on slightly distorted patterns. In order to explain this negative result, Posner and Keele presented two interpretations:

(1) During the acquisition phase the subjects needed to gather information not only (i) about the prototype but also (ii) either about the exact amount of variability among instances or about the contrastive item necessary to locate a category boundary between related categories.

(2) The training in low variability was comparable to the case of over-extension and the training in high variability, to the case of under-extension. These over-extensions and under-extensions explained the errors they obtained.

In accordance with Bruner's view (§ 3.0), the above interpretation (1) makes it clear that the knowledge of concept
involves (i) the acquisition of prototype and (ii) the information about the acceptance limit (category boundary). This acceptance limit in this thesis is called the 'judgemental limit' (see §7.4 & chap.10). In Rosch's model, the above (i) corresponds to the basic characteristic (a) presented in §3.1, and the acceptance limit (ii) above, to the characteristic (c) in §3.1 (i.e., the minimal between-category similarity). Barrett's CH relates to the acceptance limit above. Posner & Keele's experiment thus illustrated ironically that the subjects needed to abstract the two kinds of information before we can safely conclude that they have acquired a prototype. Posner & Keele's interpretation (2) indicate that the low variability training meant a great deal of similarity among the patterns presented. In this sense this training condition encouraged the subjects to induce the narrower range of acceptance limit. On the other hand, the high variability training meant a slight similarity among the patterns presented. Since the subjects induced the acceptance limits on the basis of slightly similar patterns, they conceived a broader range of acceptance limit. In this way, Posner and Keele explained the data which disconfirmed their original hypothesis.

Posner and Keele also demonstrated that people's judgement of the degree of category membership for each of the new patterns reflected the similarity relationship between prototypes and new patterns. They obtained a rank order correlation of 0.97 between distances from prototypes and error rates, which indicates that the patterns which were most distant from the respective prototypes were the most difficult to categorize. This finding ties up with the generalization about L1 children made by Greenberg & Kuczaj (1982) cited earlier: "The occurrence of under-extension and over-extension is a function of similarity." Peterson et. al (1973), however, confirmed Posner & Keele's original hypothesis above. By limiting the variation of category patterns, they found that the low variability condition promoted the formation of prototypes. Consequently the low category variability yielded a better classification
of novel instances consisting of both meaningful and meaningless (random) patterns. The difference in the two experiments seems to be due to the different treatment of data: while in Posner & Keele error rates were compared in terms of the degree of distortion, in Peterson et. al on the other hand the differing distortions were compared in terms of correct responses. The latter conservative criteria were much more rigorous than the former (1973:378). For this reason, we do not need to feel confounded by the differences between Posner & Keele (1968) and Perterson et. al (1973).

We have summarized the major experiments which led Rosch to propose PH. The summarization indicates that the three characteristics of PH mentioned above generalize the essentials of the previous studies. We have ascertained that PH involves three features:

(1) All items belonging to a category are similar in one way or another (i.e., "family resemblance" in Rosch's terminology).

(2) Each category has a prototype which the subjects abstract on the basis of instances presented. The prototype abstracted may be holistic or it may have an explicit structure consisting of various features. Rosch defined a prototype as possessing a maximal within-category similarity.

(3) Each category has its acceptance limit. Rosch defined this as the minimal between-category similarity.

§ 3.3 previous two studies on analogical reasoning

The two previous experiments on analogical reasoning are discussed here.

We have seen above that the original form of PH covers two concommitant processes: (a) the abstraction of prototype and (b) the abstraction of category limit. These two processes are reflected in the similarity judgements the subjects make (in Rosch's experiments) or the similarity measurement the experimenters compute. It is also clear now
that the subjects generalize from the acquired prototype to the novel instances on the basis of a 'similarity' recognized between the former and the latter. Because the subject's generalization is based upon similarity, the study of generalization (transfer) has been researched as "analogical reasoning". The well-known experiments are Rumelhart & Abrahamson (1973), Rips, Shoben & Smith (1973), Sternberg (1977), Brooks (1978), Medin & Schaffer (1978), and so on; from the viewpoint of artificial intelligence, there are proposals by Winston (1975), Minsky (1973), etc. which stress the importance of analogical process in learning. Two of these studies deal with superordinates and basic levels of animal names. For this reason, they are discussed here. The discussion tries to show (1)that the measurement of similarity must accord with the nature of prototype, and (2) that Rumelhart & Abrahamson's adoption of classical paradigm is useful for the L₁ situation as well as the L₂ situation.

(1)
According to Rumelhart and Abrahamson (1973), analogical reasoning can be considered as a kind of similarity judgement, taking a classical analogical paradigm of the following form: A:B :: C:D (A is to B as C is to D). Previous to their studies, Togerson (1958), Shepard (1962a,b), Kruskal (1964) etc. had developed a computational technique called multidimensional scaling (MDS) to quantify similarity judgement: (see chap. 4, 5 & 9). As Rumelhart (1978) states, he assume that one's semantic memory can be reproduced in a multi-dimensional space in which semantic features are dimensions (axes in the space) and in which each internal concept is placed along the relevant dimensional semantic features. Under this interpretation, the purpose of MDS is to reconstruct some such semantic spaces. It implies that distances in a reconstructed space reflect the degree of similarity among concepts. Rumelhart and Abrahamson (1973) also assumes that people base their judgement of similarity on psychological distances in a multidimensional space whose dimensions
are semantic features and that the judgement of similarity between concepts is reflected in psychological distances between concepts in one's long-term memory. They used Henley's MDS result on animal names (1969). Their subjects were given an analogy problem of multiple choice: \( A:B :: C:D_1, D_2, D_3, D_4 \) e.g., rat: pig: goat: (chimpanzee, cow, rabbit, sheep). They were instructed to choose one of the four alternatives so that the relationship between the first pair can be best completed in the second pair. Next they were asked to rank the order of their choices. It was hypothesized that the closer the distance (in a multidimensional space) between the first and second pair is (i.e., the interpoint distance of the first pair should be nearly equal to the distance from the C-term to the to-be-chosen item), the more would it be likely to be chosen as their first choice. They also hypothesized that the rank ordering of choices would be in reverse proportion to the distance in the space. The data confirmed their hypotheses: the observed patterns were extremely close to the predicted patterns. Another experiment gave further support to the view that analogical reasoning was based on distance relationships.

Rumelhart and Abrahamson began their analysis by assuming that the MDS result is a fair approximation of the data base in the memory from which information required for the completion of the task is retrieved. Sternberg (1977) questioned their assumption and used three alternative analyses of rated distances, cluster representation and MDS. He found that whichever method of analysis was chosen, the distances between the A & B terms was significantly correlated with the ease of analogical reasoning (1977).

Sternberg's study raises one relevant point about distances. While in Rumelhart & Abrahamson's experiment the interpoint distances in the 3-dimensional space accounted for the response distribution, Sternberg discovered however (a) that the only one dimension (humanness) contributed to the prediction of item easiness; (b) that the non-hierarchical cluster representation could account for the
variance better than the MDs configuration. In Rumelhart & Abrahamson's experiment, they adopted the interpoint distances in the 3-dimensional space. In (a) of Sternberg, the distance was considered along the single dimension of the MDs configuration. In (b), the distances were estimated on the basis of overlapping features. There are at least three kinds of similarity here each of which relates to particular norms (metrics) in the measurement of similarity. This suggests that in order to represent a specific prototype, we require the relevant metrics which correspond to the special kind of "similarity" suitable for the given prototype (see chap. 9 the relationship between metrics and processings). This seems to imply that the assumption made by Rumelhart & Abrahamson may be satisfactory. When Sternberg's analyses showed that at least his two metrics revealed the two different aspects of the data which the MDS method did not illuminate, we cannot support the assumption that the MDS configuration represents the interlexical structure in cognition (see chap. 4 & 9). The measurement of similarity must be contrived to suit the structure of prototypes (see chap. 7 & 9). Otherwise, it is not justifiable to assume the MDS result as the faithful representation of NL interlexical structure. One of the main interests of present research is to see whether the MDS representation can offer a satisfactory approximation of the NL interlexical representation (see chap. 9).

(2)
In the above experiment, Rumelhart & Abrahamson used the classical analogical paradigm which historical linguists have traditionally adopted in explaining "language change": see Bynon(1977), Stern(1931) & King(1969). The methodology adopted by Rumelhart and Abrahamson is indeed endorsed by Slobin from a viewpoint of developmental psychology:

"Developmental psycholinguistics ... shares much common ground with historical linguistics, and with the investigation of the evolution of pidgin and creole languages."
Since the acquisition of $L_2$ is often compared to the pigeonization and creolization, Slobin's view suggests that this aspect of $L_2$ acquisition is analogeous not only to $L_1$ but also to the diachronic changes of languages. Slobin argues that the underlying cognitive principles give rise to the qualitative similarities among the above three domains. Kiparsky also offers a similar view (Kiparsky, 1968).

"...we also would like to find an explanation for why language can change in the ways that they do. In that case, the reasons for assuming that simplification is a form of linguistic change become more compelling still.... Instances of morphological analogy ... are as characteristic of child language as they of historical change, although this is perhaps the most evident instance of the correspondence. ... We think of language acquisition as a process in which the child arrives at adult grammar gradually by attempting to match to the speech it hears a succession of hypotheses of an increasing order of complex hypotheses become available to the child through maturational change. For phonology this was clearly shown by Jakobson's spectacular discovery that the child learns phonemes in a largely fixed order, which is determined not externally by the order or frequency with which they are heard, but internally by their relative linguistic complexity, as reflected also in the rules governing the possible phonemic systems of the language of the world....."

(Underlying added.) Kiparsky, (1968:192-4).

In reference to Kiparsky's view, Richards (1975) mentions that "Kiparsky implies a cognitive basis for simplification, more complex rules developing as maturational development enables the learner to assimilate them": (1975:117). On the basis of King's view that "rule loss" and "rule reordering" which are major sources of language change belong to the same class of simplification, Richards
concludes as follows:

"Simplification may thus be considered as a universal learning strategy based on the extension or application of rules. Overgeneralization and analogy are instances of the same process."

Richards, (1975:118).

It is arguable whether simplification is in itself a process or the outcome of analogical process. If overgeneralization irons out the real complex TL rules, it is over-generalization that causes the complexity to look simpler (see also chap.2). From the investigator's point of view, the outcome of over-generalization or analogical process appears to look as if some simplification process is involved. But from the learner's point of view, he would not intend to simplify the complexity. Because he sees the existing rules in the light of his internalized simple rules and his speech reflects his overgeneralized simple rules, his performance merely appears to look simpler. If this analogical process or assimilation irons out the complex TL rules, simplification is not a process but the result of overgeneralization or the analogical process (see §2.2 where we have discussed the connection between simplification and assimilation). Leaving aside this minor disagreement, it is noteworthy that the analogical process is observable not only diachronically but also in synchronic changes in L₁ and L₂ acquisition processes. In this connection, it is interesting to remember Labov's 1972 data on (aw). As Bynon (1977) makes clear, Labov's data illuminates the synchronic reflection of historical change: Labov, 1972 & Bynon, 1977. Bailey's "wave theory" (1973) essentially captures the same phenomenon. Not only Dickerson (1976) but also Zobl (1982) offered the L₂ data which accord with the wave theory (see also §2.2). In this sense, the synchronic reflection of diachronic change is observable in L₂ situation (see also Koike(1983)). It may be that the internal force of analogical process may have considerable creative potentials both in L₁ and L₂ acquisition.
This section presents an operational definition of word-meaning based upon PH. This operational definition is useful in making a distinction between competence and performance. These operational definitions are presented here in preparation for the analysis in chap. 9 & 10.

We will begin our discussion with the classical analogical paradigm which has been justified by several researchers in historical linguistics, developmental psychology and applied linguistics.

In the following diagram, E denotes an object or event in \( L_1 \) and \( L_2 \) acquisitional situation. (But in the present case of translat-ability judgement (chap. 4, 6 & 10) E denotes an NL concept or sense (i.e., "the information which the name (i.e., phonetic shape) conveys to the hearer" of the same language (Ullmann, (1972:57)). \( L \) denotes a TL sign or label. Bloom & Lashley calls \( E_1 \) "the original referent situation" or "focussing event" (the latter term is from Kates(1974)). \( E_1 \) is the first linguistic event in which the original referent situation is associated with a sign. Internally, this focussing event is represented as prototype. Since a prototype can be wholistic or a set of semantic features, the focussing event can be conceived as either a single wholistic entity of experience as a whole or a collection of analytic semantic features. We regard the relationship between \( E_1 \) and \( L_1 \) is denotational (see §3.0). We have seen that according to Lyons, this assumption is appropriate in the case of \( L_1 \) children. For this reason, the relationship is written as \( D_1 \) below. But, in the case of present research about translatability judgements, it is difficult to regard the relationship as denotational; \( E_1 \) is an NL sense which is, the information the name \( 'L_1' \) conveys to himself (see above). For the sake of illustration, the present situation is also represented as \( D_1 \).

\[
\begin{array}{c}
E_1 : L_1 :: E_2 : ? \\
\hline
D_1 \quad D_2 (?)
\end{array}
\]

Diagram 1
On the other hand, \( E_2 \) denotes a novel situation or a novel NL concept which a language learner attempts to overextend or underextend his already acquired TL sign. Whether the language learner actually attempts to do so depends upon the actual context of communication. For this reason, in the above diagram the second term is written as \( ? \), and so is \( D_2 \) (?). This analogical process will arise only when there is some similarity between \( E_1 \) and \( E_2 \). Since PH says that by forming a prototype, the language learner forms an acceptance limit (the minimal between-category similarity) at the same time, he knows already the limits of application of the TL sign.

According to the above framework, the same analogical process may appear to be applicable to the infinite number of times in theory. If the learner adapts new stimuli to his own internalized prototype and maintains this type of assimilation, the learner will never come out of this infinite loop. In this case, \( D_1 \) is unchanged. The analogical process, however, can alter the content of prototype as well as the acceptance limit. (In the case of translatability judgement we use the term of judgemental limit (see \( \S 7.4 \)). The equations adopted in this thesis appears to be suitable to call it inductive limit for the focussing event and later applications of the same judgemental limit is called projective limit. These terminologies are useful in distinguishing between the process in which the learner induces a rule or concept for the first time and the process in which the learner deduces the translatability on later occasions, using the same judgemental limit). The learner is capable of altering the content of a prototype in the light of a novel experience, forming \( D_2 \), because he can go through another process of prototype formation and the altered prototype can accommodate his new experience. This assumption of recurrent analogical process has been offered by many researchers (Campbell & Wales, 1970, Clark & Clark, 1977, Kiparsky & Menn, 1977 and Bloom & Lahey, 1978). Being faced with a new experience, the learner notes that the
further differentiation and subcategorization is necessary. This process may continue until eventually he narrows or changes his prototype to fit the conditions of the TL sign used by the adult speaker of the TL. In this sense, it appears that the analogical rule extension and the process of delimiting the scope of rule application are necessary and recurrent phenomena and therefore the analogical process of prototype formation is also recurrent.

In L₁ child language development, the similarity relationship between E₁ and E₂ has been classified into two major classes of I "chained association" and II "wholistic association". The former is further divided into two: (1) "associative complex" and (2) "chain complex" (see Bloom, 1973, Bowerman, 1977 and Bloom & Lahey, 1978). Since these analogical processes are comparable to the analogical processes found in the diachronic changes of sense development, I will comment on each type of associative chains in comparison to the latter.

I Chained association

(1) **Associative complex** ( combinatory association)

As we see below, there is another associative complex. Since the two are distinct from each other, this associative complex is called 'combinatory association' here.

Werner(1948) reports the case of associative complex. His child came across a name 'qua-qua' in the context of a duck in water. ( We adopt the term of name in the sense of "the phonetic shape of the word, the sounds which make it up and also other acoustic features such as accent": Ullmann (1972:57).) This experience of Werner's child's stands for the original referent situation, forming D₁ above. On later occasions the child used 'qua-qua' in reference to a duck alone and to the water alone. The three successive speech events suggest that 'qua-qua' had two features, duck and water. In the original referent situation they are combined into one wholistic event. But later each feature is singled out. This division of components characterizes associative complex.
Associative complex is similar to what Stern calls "permutation" which he regards as one of the sources of diachronic changes of meaning. Stern defines "permutation" as "a shift in the point of view concerning a detail of a total situation"; (Stern, 1931:351). Werner's child apparently shifted his point of view and focussed on one detail of the total experience on the subsequent two occasions.

Stern implicitly assumes that a word is composed of several semantic features, and that we could more or less predict the manner in which some features are combined to create a new sense of words. For this reason, he appears to have adopted this mathematical notion of permutation. However, permutation yields more combinations than we need. For example, if a word has three features a, b & c, the total number of permutations is 15. But we only require a set of (a), (b), (c), (a&b), (a&c), (b&c) & (a&b&c) to exhaust all combinations. Stern tries to take into account the sequential order of occurrence of features; e.g., (a&b&c), (a&c&b), (b&a&c), (b&c&a), (c&a&b) & (c&b&a) are treated as different groups. This is unnecessary in our situation. In combination, the order of occurrence is not taken into account, and the above 6 groups represent the same combination of (a,b & c). We are only interested in which combination of features the language learner uses in his future speech event. Combination is more suitable for this purpose than permutation is. For this reason, we will call this type of associative complex 'combinatory association' here.

(2) Chain complex

Bloom and Lahey illustrate chain complex in reference to Vygotsky (1962). The child in their example uses the word 'water', when he is in the bath; then, he does so when he is holding a drinking glass with some water in it; later, he refers to a pane of glass window by using the same word. In this example, the child uses the same sign for the three different referents, E¹, E₂ & E₃. (see Diagram 1). The chain complex is defined as the case in which there is some similarity between E¹ & E₂ and
between \( E_2 \) & \( E_3 \), but nothing in common between \( E_1 \) & \( E_3 \). So, the characteristic of chain complex is that there is a shift in the criteria of reference and that only two consecutive speech events have some similarity.

We may relate this chain complex to "adequation" (Stern (1931:381) and Waldron, (1979:132)). Waldron explains:

"Adequation is the change of meaning which results when the original reason for the choice of a particular name gets forgotten or disregarded and some new aspect of the object is seized by the mind as the meaning of the word."

Waldron takes Stern's example 'horn'. We can compare the case of 'horn' with the child's 'water'.

Historical change \( L_1 \) chain complex
(a)'animal's horn' (a)'water in the bath'
(b)'animal's horn' (b)'water in the drinking glass'
used for music'
(c)'musical instrument' (c)'a pane of glass window'
made from animal's horn'
(d)'instrument for producing a certain kind of sound'

According to Waldron, the change from (b) to (c) illustrates adequation, since there is a shift of criteria of reference. He regards the other changes (a to b and c to d) as non-linguistic in origin, suggesting that the technological progression of manufacturing musical instruments has brought about changes of meaning. Because "substitution" is defined as non-linguistic in origin, Waldron regards the other changes as the instances of "substitutions". But in the case of the child's chain complex, the changes from (a) to (b) as well as from (b) to (c) involve a shift of criterion and the instances of adequation.

While adequation evolves over time, a chain complex occurs in a short span of time. So, the rapid shift of criteria of the latter makes us wonder whether it is merely a whimsical aspect of a child's mind. Bowerman argues
that chain complex is highly unlikely to exist in reality (1977:246), saying that the rapid shift implies no properly defined intension in the mind of a child. This unstable shift of criterion might reflect the total absence or only defective presence of a mental representation for words in the mind of a child.

In comparison with the full adult lexical meaning, the child's prototype seems defective, but we must not ignore that in both cases the prototypes are equally the product of the selfsame formation process. For this reason, we assume here that as in combinatory association, all relevant features are abstracted in the original referent situation of $E_1$ (see Diagram 1). This assumption explains the capricious shift as consistent. The difference between chain complex and combinatory association is that in the former the shift of criteria of reference is more striking than in the latter, making chain complex look like childish inconsistency. The chain complex could be a means by which an infant informs his caretaker of his partial knowledge by highlighting one detail of his total experience he had at $E_1$.

(3) **Associative complex**

This associative complex refers to the cases in which "the different instances are all similar in some way to the first instance, but are not necessarily similar to each other"; Bloom & Lahey, (1978:122)). They give an example from Bowerman's data. The child first uses the word 'gi' when she is bouncing on a spring horse; then, the same word is used for horse, a man driving a tractor, a boy riding a tricycle, horses on TV, etc.

But in some cases, associative complex is based upon a few criterial features: see Lewis's example of 'fafa' (Bloom & Lahey, (1978:123-4)).

This associative complex may be similar to "transfer" which occurs in diachronic changes of meaning. According to Waldron, transfer is "a kind of metaphor", but "transfers are caused by the subliminal perception of similarities and are not the result of any conscious metaphorical intention. The word brick, for example, may be
unreflectingly used of objects of brick-like shape."; (Waldron, (1979:132)). Both Stern and Waldron make a distinction between metaphor and transfer on the basis of the presence or absence of intention. We are not absolutely sure whether the above L₁ examples are conscious metaphors or unconscious transfers. But the child's rapid shift of referents for 'gi' or 'fa-fa' involves 'some subliminal perception of similarities' and they illustrate synchronic examples of diachronic transfers.

II Wholistic association

(4) Wholistic association (superordinate association)

Wholistic association relates to the presence of superordinate semantic features in prototypes. This association has been documented in many diary studies (see Clark, 1973, Bowerman, 1977). One example often cited is that the child used the same word 'tee' for dogs, cats, horses and sheep covering all the four-legged animals he had come across. I have observed several cases among Japanese infants. One infant used 'boo-boo' for cars, trains, bicycles, trams and baby carriages covering a category of vehicles he has seen. In some cases, a particular dog's name was used by the child to refer to any dog. All men except for his own father were called by his own brother's name. These examples seem to indicate that a very young child can abstract a superordinate feature for a category.

We regard wholistic association as separate from associative complex. This is because the abstraction of a superordinate feature is more like the abstraction of a central tendency in the psychological experiments discussed above. Whereas, in the latter case of associative complex, any feature can initiate associative complex in Eᵢ, as long as a feature in each Eᵢ has some similarity with E₁.

The wholistic association relates to "transfer" which is one of the sources for diachronic changes of meanings. Waldron's definition accords with the usual characterization of wholistic association and requires no further comment.
So far we have examined the four kinds of overextension treating them as prime cases of lexical performance. We have also seen that these four types which occur synchronically can be usefully compared with diachronic changes of meaning. We are tempted to say that we are endowed with these creative capabilities which are reflected in the history of sense development.

Operational definition of word-meaning
In view of the four kinds of over-extension, we may say that $E_1$ (Diagram 1) possesses all relevant features some of which are focussed in the subsequent occasions $E_j$. From this, we can get a very simple, but convenient definition of word-meaning, following Rosch's maximal within-category similarity. At the same time, on the basis of her minimal between-category similarity, we can distinguish competence from performance also operationally.

We will use "distributive lattices" as a framework except for the first illustration for three reasons. (1) Ullmann argues that there is a reciprocal and reversible relationship between sound and sense as well as between name and sense: "It is this reciprocal and reversible relationship between sound and sense which I propose to call the 'meaning' of the word" (Ullmann, 1972:57). This reciprocal and reversible quality is best expressed by distributive lattices, as Piaget makes clear. (2) According to Yabuuchi (1982), Piaget's INRC which is based upon distributive lattices is compatible with the general form of analogical reasoning. (3) The use of distributive lattices connects the present definition of word-meaning with our discussion of processing in chap. 9. It seems to me that this is the only way in which the mathematical side of MDS can be usefully related with psychological or linguistic discussion. Additionally, Piaget's accommodation and assimilation which Corder (1978b) regards as operative in the $L_2$ acquisitional process may be revealed, but this is largely a topic of future research.

(As Putnum (1975) makes clear, natural languages are not
particularly suited for an explicit treatment. For this reason, Putnum proposes the notion of stereotypes as more tenable. I do not present any discussion concerning the relationship between Putnum's stereotypes and the psychological notion of prototypes discussed above. This is also a topic of future research.)

The term 'feature' is used here in a very broad sense to indicate any organizational schema, including appearance (eg., the transparency of water in the above examples), texture, posture or some wholistically conceived entity (duck in water above). We are only concerned with operational definitions, not with a proper linguistic analysis. In the above, we have decided to start with the simple criterion of all or some. If all features contained in a prototype the language learner abstracted at $E_1$ are verified, we may suppose that this is closer to the prototypical situation (see chap.9) and therefore closer to our intuitive understanding of 'competence'. Likewise, if some features are verified in the immediate context of $E_2$, it relates to our understanding of 'performance. Let us suppose that there are three features ($b_1, b_2, b_3$) in the abstracted prototype. The language learner would verify as follows:

<table>
<thead>
<tr>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$\land$</th>
<th>$\lor$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

Table 1

Rows 1 - 7 exhaust all combinations of three features which we have dealt with the combinatorial association (1). Since we have defined competence as all features being verified as true, we can see that the first row satisfies this
definition. Row 1 represents \( b_1 \land b_2 \land b_3 \), since all the three features are verified. More generally, we can represent this set intersection as

\[
(1) \quad \bigcap_{i \in I} b_i \quad \text{or} \quad \prod_{i \in I} b_i \quad \text{where} \quad \forall_i \in I
\]

(\( I \) represents a category and \( b_i \), constituent features. This defines the operational intentional definition. But as we see in §9.2, we can use this as the operational extensional definition. In this case, \( b_i \) represents the members of category \( I \).)

The unification shown in Table 1 covers the four kinds of lexical performance discussed in this section. If one of the three features \((b_1, b_2, b_3)\) stands for a super-ordinate feature, this illustrates the superordinate association (3) and rows 5 - 7 account for this lexical performance. The same explanation applies in the case of associative complex. Since this involves a few (not all) criterial features, rows 2 - 7 explain the lexical performance based upon the associative complex. The combinatory association relates to rows 2 - 7. The chain complex, in the way we have interpreted it, relates to rows 2 - 7; if \( E_1 \) initiates row 5, \( E_2 \) may initiate row 6: and \( E_3 \) initiate row 7. The three successive \( E_1, E_2 \) and \( E_3 \) would show that there is some similarity between \( E_1 \) & \( E_2 \) and between \( E_2 \) & \( E_3 \), but that there is no clear similarity between \( E_1 \) and \( E_3 \). This characteristic of chain complex is explained by rows 5 - 7. Thus, if the performance is based upon one feature, rows 5 - 7 explain the process involved in this lexical performance. If the lexical performance is based upon two features at the same time, rows 2 - 4 account for this. In this way, we can recognize that rows 2 - 7 represent our definition of performance in which not all but some features are used by the language learner. Rows 2 - 7 stand for \( b_1 \cup b_2 \cup b_3 \). We can represent this in a more general form as our operational definition of performance:

\[
(2) \quad \bigcup_{i \in I} b_i \quad \text{where} \quad \forall_i \in I
\]
So far we have argued that competence is represented by expression (1) and performance by expression (2). We will examine the appropriateness of these expressions in three ways. (A) The representation of performance accords with Tversky's formula for metaphorical uses of language. (B) The operational distinction of competence vs performance is in keeping with Ullmann's conception of word-meaning, i.e., the reversible and reciprocal relationship between sound and sense. (C) The two expressions (1) & (2) corresponds to Rosch's maximal within-category similarity and her minimal between-category similarity respectively.

(A) Tversky (1977) attempts to account for the role of similarity in the metaphorical use of language as well as other similarity judgements involving visual or conceptual materials. Tversky proposes the following formula to define analogical thinking.

\[ S(a, b) = \theta f(A \cap B) - \alpha f(A-B) - \beta f(B-A) \]

\( S(a, b) \): the similarity function from a to b.
\( f \): the measure of salience of shared features \((A \cap B)\) and distinctive features of \((A-B)\) and \((B-A)\).
\( (A \cap B) \): the common features between A & B.
\( (A-B) \): the features distinctive to A.
\( (B-A) \): the features distinctive to B.
\( \theta, \alpha, \beta \): the parameters that reflect the importance of the shared & distinctive features.
The saliency may vary, depending on whether the function deals with the common features (A B), the distinctive feature of A (A-B) or the distinctive feature of B (B-A). For this reason Ortony (1979) modified the above function as follow:

\[ S(a,b) = \theta f^B(A B) - f^A(A-B) - \beta f^B(B-A) \]

According to Tversky, there are two principle factors to determine the salience of a feature: intensive and diagnostic. They are explained as follow:

"Intensive factors: the brightness of a light, the loudness of a tone, the size of a letter, the frequency of an item, the clarity of a picture, or the vividness of an image.

Diagnostic factors: the classificatory significance of features, that is, the importance or prevalence of the classifications that are based on these features."

Tversky, (1977: 342)

Both intensive and diagnostic factors relate to what we have loosely called 'features' contained in a prototype. The degree of intensity and clarity of constituent features of prototypes may vary, depending upon the immediate context of situation E_j (see Diagram 1). In this sense, Tversky's formula takes into account performance factors.

But the major determinant of saliency does not originate from extraneous factors, but from the subjective/psychological fact of which particular feature the subject's attention is drawn to in the immediate context of situation. It may be largely due to the contingency of particular speech events that determines the direction or focus of our attention. In this sense, it is meaningful to consider Tversky's or Ortony's equations, by keeping
all attentional factors constant (say unity). Then, we discover that his equation is reduced to the sum set which is equivalent of our eq.2.

\[ a, b, \theta = 1 \]

\[ S(a,b) = (A \land B) - (A \land \neg B) - (B - A) \]

\[ A - B = \{ x \mid x \in A \land x \notin B \} \]

\[ = (x \mid x \in A) \land (x \mid x \notin B) \]

\[ = A \land B^c \]

\[ = (A \land B) - (A \land B^c) - (B \land A^c) \]

\[ = (A \land B) \cup (A \land B^c)^c \cup (B \land A^c)^c \]

\[ = (A \land B) \cup (A^c \cup B) \cup (B^c \cup A) \]

\[ = (A \cup B) \]

\[ B - A = \{ z \mid z \in B \land z \notin A \} \]

This suggests that Tversky's or Ortony's equation is based upon the set sum of features some of which are highlighted by attentional factors and that three weights \( a, b \) \& \( \theta \) adjust the amount of saliency. In this sense, Tversky's and Ortony's equations essentially expresses the same view as eq.2 above; i.e., not all but some features are actively used in performance. The metaphorical use of language properly belong to performance. It is significant that the present definition (expression 2) corresponds to Tversky's (or Ortony's) formula of metaphorical use of language in its essence.

(B) We have already noted Ullmann's view of word-meaning. He regards the reciprocal and reversible relationship between sound and sense as constituting the meaning of word. The above definitions of eq.1&2 satisfy Ullmann's view of word-meaning.

The examination relates to our use of distributive lattices. This allows us to use Piaget's INRC and it is also convenient for the discussion of processing in chap.9. Piaget's INRC stands for identity (I), negation (N), reciprocal (R) and correlative (C) operations. The relationship among them is diagramed as follows:
Among the four operations of INRC, correlative closely relates to a reciprocal and reversible relationship:

\[ C(\overline{E}) = \overline{R}(\overline{E}) \]

Ullmann proposes a reciprocal and reversible relationship to capture that a sound we hear reversibly calls forth to its mental representation and that a sound relates to sense reciprocally. Eq.(5) indicates that correlative involves reciprocal operation internally and the above schematic diagram indicates that mental representations (senses) correlative relates to reality (sounds and letters which physically embody senses).

\[ b_1 \wedge b_2 \wedge b_3 \iff b_1 \vee b_2 \vee b_3 \]

The above two equations (6) & (7) indicate that even if some features are available and attended in reality, all features are accessible in competence (see chap.5&9). Eq.(6) shows that when the language learner receives a
lexical event, he would verify it, based upon one of the six verification processes shown in Table 1 and the correlative operation restores all relevant features in his mental representation as his linguistic competence. Eq.(7) says the opposite: the language learner produces a lexical event on the basis of competence, but, some features are used, as eq.(7) indicates.

Additionally, it may be convenient to distinguish two situation: prototypical situation and non-prototypical situation. All original referent situations are prototypical by definition. But as Table 1 shows, we must be prepared to accommodate the possibility of row 1 in which a prototypical situation is reproduced in the subsequent occasions of Ej. The prototypical situation (during which a prototype is formed) is reproduced as it was and all relevant features are available for verification. (As we see chap.5 we can make a distinction of availability and accessibility in memory as well. Once a prototype is abstracted, all features comprising the prototype is accessible, but some features are more readily available for immediate use.)

\[(C)\] Here we will examine eq.1&2 in view of Rosch's proposal of maximal within-category similarity and minimal between-category similarity. This time we regard eq.1 & 2 as specifying the membership of a category (i.e., words belonging to a superordinate word) instead of viewing \(b_i\) as a collection of features. This is because Rosch's approach is based upon memberships and the intensional definition of eq.1 & 2 in fact conforms to the extensional definition (see §9.2). The latter treatment of eq.(1) & (2) is in fact the standard method (see Takeuti, 1975).

We regard eq.1 & 2 as specifying the membership of a category. Rosch's framework involves more than two categories. For this reason, it is better to specify the range of \(b_i\):

We assume \( \{ b_i \mid i \in I \} \subset B \).

If \( \forall_1 \in I \ ( b \geq b_i ) \),

\( \sum_{i \in I} b_i \) is the "least element" in \( b \).
If \( \forall i \in I \ ( b_i \geq b ) \),
\( \prod_{i \in I} b_i \) is the "greatest element" in \( b \).

We can schematize these least and greatest elements as follows:

\[
\begin{array}{c}
\text{Category 1} \quad \text{Category 2} \quad \text{Category 3} \\
\prod_{i \in I} b_i \quad (b_i \geq b) \quad \sum_{i \in I} b_i \quad (b \leq b_i)
\end{array}
\]

Diagram 3

In Diagram 3, Category 1 is in the "lower bound" of Category 2; and Category 3 is in the "upper bound" of Category 2. Since \( \prod_{i \in I} b_i \) is the maximal element in \( b \), it is the greatest element in the lower bound (glb)(i.e., Category 1 in Diagram 3). Since the framework we are adopting is distributive lattices, glb is equal to the minimal element belonging to Category 2. This suggests that this glb characterizes the best element of Category 2, since it has all features the other elements of Category 2 process. That is, glb characterizes the maximal within-category similarity. (We must bear in mind that although we are adopting here an extentional definition, it is known that Zorn's lemma proves the existence of underlying intensional properties which confirm the extensional definition.)

The case of \( \sum_{i \in I} b_i \) is straightforward. It is the least element in the upper bound (lub) of Category 3. This is the smallest element in the upper bound and is equal to the greatest element of Category 2. Thus, this lub suggests the minimal between-category similarity, since it is the greatest element of category 2 and hence the poorest example of Category 2 as we are adopting distributive lattices. The above illustration is somewhat over-simplified, but it is sufficient for the present purpose (for further discussion, see §9.2 and §9.3.2 (2)).
We have used 'distributive lattices' with Piaget's accommodation and assimilation in mind, in characterizing word-meaning. We have seen that (A) the present definition of performance accord with Tversky's definition, (B) Ullmann's proposal of word-meaning fits in with the present definition and (C) Rosch's notional definition of maximal within-category similarity and minimal between-category similarity can be made explicit within the framework of distributive lattices.
Chapter 4: pilot study (1); follow-up study of Kellerman (1978)

§4.0 purpose

The methodology of this pilot study follows in part Kellerman (1977, 1978 & 1979). However, four tentative approaches which are not considered by Kellerman are incorporated. (1) Kellerman regards one of the two or three dimensions yielded by multi-dimensional scaling (MDS) as representing the continuum of "core" and "non-core". This single dimension is seen as predicting a learner's judgement of translatability in a straightforward manner. That is, the closer a NL sentence is to the "core", the more readily it is accepted as translatable by the learner. The degree of acceptance falls along the single dimension named by Kellerman as the "core-noncore" dimension. Thus, the effect of a learner's processing upon his judgement is not considered by Kellerman's study. Here it is postulated that task-oriented processing intervenes between the NL inter-lexical structure and the judgement of translatability. Garner has suggested two kinds of highly general processing with special emphasis on multi-dimensional stimuli (Garner, 1976, 1974a and 1974b). These are called respectively "integral" (holistic) and "separable" (analytic). The latter has been further classified into two types: separable and "asymmetric separable" (ibid.). These distinctions are explained below. (Kellerman's analysis suggests that his subjects paid attention, either consciously or unconsciously, to one specific dimension. This implies that his subjects had adopted asymmetric separable processing in making their judgements.) The distinctions mentioned here have been experimentally demonstrated in the areas of perceptual processing, concept learning and choice behaviour (see the experiments cited in Garner 1976). More recently, it has been demonstrated that the distinctions are related to cognitive development (Shepp & Swartz, 1976; Shepp, 1978) and also that the specific processing adopted by a subject depends on the nature of the experimental task: Foard & Nelson, 1984; Smith & Nelson,
1984. The purpose of this pilot study is to see whether Kellerman's result can be observed even when we incorporate the intervention of processing into the analysis of data.

Kellerman's study deals with one verb, but when his sentences are translated into Japanese, four verbs are required (see below). Analysis I replicates Kellerman's data treatment in the sense that the MDS result is regarded as offering an adequate representation of the NL interlexical structure. Based on this assumption, the judgement data were then analyzed according to Kellerman's implicit assumption of asymmetric separable processing. It is shown that only 2 out of the 4 verbs support Kellerman's position.

(2) In Kellerman's data drawn from 10 groups of subjects (ranging from 2nd year secondary school students to 2nd year university students), there were always one or two items which were accepted by almost all subjects: 100% in 7 groups; 96%, 97% and 83% in the other 3 groups. This suggests strongly that there are one or two clear cases from which their judgements about the unclear cases might have been derived. It seems probable that, as a result of learning, the subjects are sure of the translatability of one definite NL item, and that they base their judgements of the rest of the items upon this confidence. Then, they intuit the degree of similarity between the 'sure' item and other items (for this process of analogy, see (4) below).

Analysis II incorporates the effect of the 'sure' item(s) on the judgement process. As in analysis I, analysis II assumes that the MDS result represents a satisfactory NL interlexical structure. Thus, the subject's 'feel' of similarity reflecting his analogical process is numerically represented as the distances in the MDS configuration. Analysis II considers the role of the 'sure' item(s) in relation to the two kinds of processing: asymmetric separable and integral.
In the course of discussion Corder's definition of accommodation and assimilation is introduced only in relation to Analysis II (see chap. 2). Assimilation is used here to mean 'conceiving a novel event in the light of previous experience' (e.g., NL transfer is to conceive TL expressions in the light of NL). Assimilation is normally regarded as involving no change in the established internal structure, since it makes use of any previously established knowledge of either TL, NL or interlanguage, but Ausubel's view is different (see chap. 2). Whereas accommodation involves change; in order to accommodate a novel experience, either TL or NL, a part of the previously established knowledge goes through some change.

Analogical process also relies upon some such previously established knowledge. It stresses the operation of similarity, but the previously established knowledge is a basis upon which the 'feel' of similarity is intuited. Without any previously established knowledge, the 'feel' of similarity does not appear to arise.

The idea of 'sure' item, if it is properly specified, has some theoretical significance. The 'sure' item on the part of a learner represents a fragment of previously established interlanguage (i.e. knowledge). If we do not have any means of identifying a learner's 'sure' item, we do not have any idea of the internal basis upon which analogical process and NL assimilation is carried out on the part of a learner.

(3) The subject's feel of similarity between the 'sure' item and other items plays a major role according to this interpretation (2). This is different from Kellerman's position. While the present interpretation focuses on the importance of the inference made from this specific (and confident) item, Kellerman's study stresses the more general notion of 'feel' of "core". As we have seen, Kellerman gives one of the dimensions in the MDS configuration the name of "core and non-core dimension", by regarding the MDS configuration as representing the
learner's subjective NL lexical structure faithfully. This suggests that a part of the learner's lexical knowledge includes cross-linguistic awareness of 'core'. However, as Kellerman rightly admits (1979), his nomenclature is essentially putative. Contrary to what is sometimes assumed, the technique of MDS does not designate the names of dimensions. Although the user of MDS is encouraged to name the dimensions by looking at the configuration of stimulus points corresponding to each input item (in the spatial representation), the computational technique of MDS is purely mathematical and does not justify putative nomenclature. In order for us to name the dimension legitimately (particularly the polar continuum of core and non-core), cross-linguistic analysis is strictly required. For this reason, in this pilot study such naming is avoided apart from a few occasions in which Kellerman's naming is considered.

Rather, the strategy of paired comparison is postulated. This assumption is in one important respect as general as the assumption of 'core'. The conceptual comparison of two items appears to be a primitive and therefore general operation. As we have seen in chapters 2 and 3, the analogical process is observed both among second language learners and among first language children. The process is manifested variously as overgeneralization including overextension, underextension, NL transfer and lexical simplification in pidgin languages and interlanguages. The paired comparison can be regarded as the basic unit of such analogical processes, since comparing two items at a time is the most primitive mode of comparison. In this sense, the generality comparable to 'core' might be maintained (see (4) below). Here, the subject is assumed to compare his 'sure' item with another item on the list. The pair can be compared in two ways. (1) One pair is compared one at a time sequentially in the order of the intuitive 'feel' of similarity. The closer a given item is to the 'sure' item, the earlier it is compared with the 'sure' item. (2) Or alternatively the paired comparison might be carried out among several pairs
simultaneously. This pilot study does not make any specific assumption about which type of paired comparison is made by the subject. As we see in chap. 10, the above two types occur, but for the time being we try only to see whether the paired comparison occurs at all. In pilot study (4), free recall experiments are run to see whether the paired comparison accounts for the process of judging translatability. The results support this observation.

Under the above assumption of paired comparison, Thurstone's (and Torgerson's) classical method of measuring the 'feel' of semantic similarity is used. The Thurstonian method modified by Torgerson (1958) is used here to obtain numerical approximations for the 'feel'. The method of measuring the 'feel' of similarity simulates the process of paired comparison, because the equation appears to be contrived so that the values along the continuum (the linear scale) of similarity reflect the paired comparison which we assume the subject is adopting in his judgemental process. Pilot study (5) examines whether this application has some plausibility - this experiment is necessary to justify why this method (originally designed for psychophysical measurement) is used here.

MDS requires as input the ranked data of all possible combinations of pairs. In eliciting the input data, I had another purpose in mind: measuring the 'feel' of similarity between the paired NL sentences. Using Torgerson's equation, the same data was converted into the pairwise measures of similarity between NL sentences. Analysis III considers relationships between pairwise similarity measures and translatability judgements.

(4) The learner's reliance on the 'feel' of similarity, either between two items or among several items, is in the present study interpreted equally as belonging to the analogical process. We have already seen in chapters 2 and 3 the importance of this strategy in explaining the well-known phenomena of overgeneralization, NL transfer and lexical simplification. By postulating the analogical
process as an underlying mechanism, the three kinds of phenomena are seen as derived from the self-same cognitive activity. Furthermore, the role of analogy is well recognized in researches into first child language and previous experiments demonstrating the process of analogy are touched on in chap. 3.

This pilot study regards the process of analogy as justifying the method of analysis which relates to the following questions: (1) whether it is legitimate to rank the measured 'feel' of similarity so as to see a correlation with the judgement data; (2) whether it is meaningful to convert the ratings of paired sentences (the method derived from Torgerson's) into values along a scale of similarity continuum; (3) whether it is justified to rank the distances derived by MDS in order to see the correlation with the judgement data. These methods are only justifiable if the subject intuits similarity (i.e., the analogical process occurs) and his judgement is based upon it.

The two kinds of processing considered here relate merely to the issue of how the subject organizes the NL lexical structure either from the unanalytical mode of global grasp (integral) or from the more analytical mode (separable). They do not directly relate to the process of intuiting similarity. What triggers the process of intuiting similarity is apparently different from the issue of how the subject processes the NL knowledge. Thus, it seems to be plausible to make a distinction between the process of analogy and processing. In this pilot study and also in later chapters, the 'feel' of similarity is assumed to be the result of analogical process. It is manifested in the 'feel' of similarity. In this sense the measurement of similarity is regarded as crucial. This is the reason why pilot study (4) examines whether the application of Torgerson's equation has some psychological plausibility. However, in this pilot study, the equation adopted is the simplest of all (condition IV). This pilot study is a preliminary attempt in this direction.
§4.1 summary and comment: Kellerman's study

This section summarises Kellerman's research method with special emphasis on his use of MDS. It is discussed below that the MDS method does not require an investigator to make any intuitive assessment concerning a learner's view about his NL. It is important to bear in mind that the ways in which the second language learner intuits the structure of his NL may be different from an investigator's view. Kellerman's methodology is learner-oriented. It does not require or depend upon his observation about a learner's NL. The use of non-metric MDS makes this learner-oriented paradigm possible. Input data for MDS are gathered directly from learners. Being non-metric, MDS only requires ordinal data reflecting similarity judgements about the paired NL items. The input data in Kellerman's experiment were the impressionistic rankings of paired NL sentences made by the subjects (see, Kellerman, 1978). In the other experiment the subjects sorted cards on which NL sentences were written. The ways in which cards were sorted give some measure of the degree of similarity recognized by the subjects. MDS claims to abstract the latent criterion on which the subjects' 'feel' of similarity is based. Thus, the interlexical NL structure represented by MDS is learner-based. Since it reflects a learner's subjective 'feel', Fillenbaum and Rapoport regard MDS results as recapturing the structures in the "subjective lexicon" (1971) and others, as "a map of mental states" (e.g., Clark, 1983). To sum up, Kellermans's application of MDS among second language learners reveals learner-generated intuitional data about their NL without involving the investigator's opinions about the comparison between NL and TL.

Kellerman presents two MDS results: the one derived from the 9 sentences which are also used as the materials for the judgement test and the other MDS result obtained from 17 sentences which include the original 9 (1978). Instead of using the former MDS result where all 9 items in the judgement test appear in the MDS map, Kellerman presents
correlations (in terms of Kendall's measure of disarray (S)) between the learners' intuition about the "transferability" (translatability) of the 9 sentences and the "core dimension" of the latter MDS result. According to the result (1977), the "core" dimension in the former MDS does not appear to show significant correlation in at least three groups (see below).

But the MDS configuration reveals two clusters. When the judgement data are classified according to these clusters, the correlations improve exceedingly. The importance of clusters in the MDS configuration will be discussed below. All 8 groups in Kellerman show no significant differences in terms of the rank order of acceptance regarding the 9 sentences. However two more groups are then added. This time, the 17 sentences are used for the translatability judgement test (1978). The acceptance order of the 17 items correlates highly with the rank order of the 17 items in the "core dimension"; \( \rho = 0.837 \) in the 2-dimensional solution and \( \rho = 0.639 \) in the 3-dimensional solution. Moreover, two further groups of Dutch learners of German judge the translatability of the original 9 sentences. In this third case, the order of their acceptance correlates with the rank order of the 9 items depicted by the MDS result for the 17 items. These correlations are again shown in terms of Kendall's measure of disarray (S). Kellerman claims that the "core" dimension has "predictive power" for the learners' intuitive transferability.

Three minor comments can be made in relation to Kellerman's data treatment. The rank order correlation is in most cases considered in terms of Kendall's S (see below) rather than Kendall's correlation coefficient \( \tau \), although Spearman's correlation coefficient \( \rho \) is used a few times. S is the measure in the process of computing \( \tau \):

\[
\tau = \frac{S}{\frac{1}{2} N(N-1)}
\]

No explanation is given for presenting S rather than
In the cases where $S$ is presented rather than Spearman's $\rho$, according to the information supplied, we can calculate the values of $\rho$; sometimes the values of $\rho$ are not statistically significant. In the following table, a part of Kellerman's Table 12 is reproduced (1978). $\rho$ is computed only in the cases where the lowest $S(17)$ and the highest $S(24)$ were reported in his Table 12:

<table>
<thead>
<tr>
<th></th>
<th>$S$</th>
<th>sig</th>
<th>(level of significance: $\alpha = 0.05$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KU2</td>
<td>17</td>
<td>.05</td>
<td>0.5833 (0.05&lt;p) N.S.</td>
</tr>
<tr>
<td>KU1</td>
<td>19</td>
<td>&lt;.05</td>
<td></td>
</tr>
<tr>
<td>KU0</td>
<td>17</td>
<td>.05</td>
<td>0.55 (0.05&lt;p) N.S.</td>
</tr>
<tr>
<td>A6</td>
<td>21</td>
<td>&lt;.05</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>21</td>
<td>&lt;.05</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>21</td>
<td>&lt;.05</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>24</td>
<td>&lt;.01</td>
<td>0.5167 (0.05&lt;p) N.S.</td>
</tr>
<tr>
<td>A2</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3</td>
<td>22</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>22</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

Kellerman, (1978:20)

**Table 1**

Kellerman interprets the above result as indicating that there is "generally strong" correlation between the rank orders of translatability judgement and the rank orders in the "core" dimension. However, if the probability $p$ yielded is larger than $\alpha$, the null hypothesis ($H_0$: there is no significant association between the two variables) must be accepted. Thus, in the above, according to the probabilities associated with the occurrence under $H_0$ (coded as sig in Kellerman's Table), the three groups show no significant rank order correlation ($\rho = 0.01$) with the "core" dimension.

It is to be noted that the better correlation is observed when the judgement data about the 9 sentences are compared, not with the MDS result for the same 9 sentences, but with the MDS result produced by the 17 sentences (see,
Kellerman's Table 12, 1978:20). In my limited experience as a user of MDS, it is often the case that when we change the number of input items, the configuration changes and therefore the rank order of the input items along the dimensions alters accordingly. This is precisely the case in Kellerman. His original MDS with the 9 items presents the rank order of 7, 6, 9, 3, 1, 4, 5, 8, and 2 along the "core" dimension (i.e., dimension 3 in the 3-dimensional solution). On the other hand, the MDS result with the 17 items presents the order of 7, 3, 6, 9, 5, 1, 8, 4 and 2 along the "core" dimension in the 3-dimensional solution. Apart from the items 7 & 2 the order of the other items is altered. It is arguable whether the acceptance order of the original 9 sentences can be legitimately compared to the MDS derived from the 17 items. Furthermore, in choosing a 3-dimensional solution, no explanation is supplied for his choice of one graph, out of three, in which dimension 1 is plotted against dimension 3. In the case of the 3-dimensional solution, there are two other graphs: dimension 1 plotted against dimension 2 and dimension 2 plotted against dimension 3. In addition, it is not clear why in reporting the MDS result for the 9 sentences the 3-dimensional solution is chosen rather than the 2-dimensional solution. These are merely the minor points, but they are not explained sufficiently by Kellerman. We have seen that, as a learner-oriented approach, Kellerman's method is an important development in the investigation of the process of learning and acquisition. In comparison with the more familiar method of performance (product) analysis, his approach is more process-oriented.

§ 4.2 the operational definitions of three types of processing

Although Kellerman's application of MDS is insightful in representing a learner-oriented NL inter-lexical structure, it is arguable whether the learner has actually made his translatability judgement along the single dimension of "core" and "non-core". There are other possibilities.
For instance, the learner being aware of various semantic features, his judgement might be derived from his overall assessment of two or more dimensions. The other possible interpretations become important when we see the translatability judgement as the interactive result of processing and the NL interlexical structure.

\[ \text{Diagram 1} \]

The types of processing considered are (a) integral (holistic), (b) asymmetric separable (analytic) and (c) separable (analytic) processing. Although separable processing is not properly dealt with here, the following notional definitions refer to these three processes (see Garner, 1976): Garner's definitions are cited and commented in §9.2.

(a) **Integral (holistic) processing** involves a global assessment. The subject processes the item holistically as an unanalysed whole. The unanalitical 'feel' of similarity and the perception of color illustrate this processing. The perception of color involves brightness, saturation & chroma, but we perceive color without analyzing these overtly. This processing can be explained in reference to multi-dimensional stimuli such as word concepts, as in the present study.

Let us take a hypothetical situation and assume that the following diagram is the MDS result we obtained. If the subject adopts integral processing, each of dimension (DIM) 1 and DIM 2 is not relevant on its own, since he does not analyze the stimuli A, B & C. The judgement under the influence of integral processing is derived from the interpoint distances: the distance between A & B, the distance between B & C and the distance between C & A. The closer the interpoint
distance between the pairs is, the more similar the pair is judged to be.

In our situation, MDS computes the interpoint distances for each pair. If we assume that the subject has adopted an integral processing, the experimenter only needs to see whether interpoint distances can correlate with the judgement data.

(b) Asymmetric separable (analytic) processing involves one strongly favoured dimension. All relevant dimensions are separately noted by the subject, but they are not equally favoured. One of the dimensions is, consciously or unconsciously, preferred and the subject's attention is concentrated upon it. In the following diagram, if dimension 1 is favoured and the subject adopts a similarity-based strategy (i.e., the analogical process), his performance would reflect the similarity order of C, A, B and D or its reverse. Likewise, if dimension 2 is favoured, his performance would reflect the similarity order of A, B, C, and D or its reverse.
The MDS computes the exact values for the respective items along the dimensions. Once we assume that the MDS result offers a reasonable approximation of the NL interlexical structure, the experimenter needs only to see the correlation between the observed judgement data and the values along each dimension which correspond to the NL sentences. The correlation coefficients we will obtain determine whether the asymmetric separable processing was adopted by the subjects and which dimension was used by the subjects in the judgemental process.

(c) Separable (analytic) processing
As in the asymmetric separable processing, it involves relevant dimensions which can be selectively (separately) attended to. The selective attention to each of the relevant dimensions presupposes the presence of some well-defined structure in the mind of a learner. It also presupposes his ability to note all relevant dimensions separately (see Clark’s SFH discussed in chap. 3). Thus, as developmental psychologists demonstrate, this processing is more sophisticated than the two processings mentioned above: see Shepp & Swartz, 1976; Smith & Kemler, 1977 & 1978; Shepp, 1978.

The learner must be able to note the constituent dimensions which comprise the well-defined structure. The above cases of integral and asymmetric separable processings, the manner in which the two processings influence the judge-mental process is straightforward. However, purely separable processing cannot be in theory determined merely by looking at the correlation between the MDS and the judgement data. As Garner makes clear (1976), it depends upon whether the assumption of "city-block" metric or Enclid metric yields a better fit (see chapter 9). However, by applying Shepp’s notion of correlated and orthogonal stimulus pairs, it is possible to make some prediction. Let us suppose the following hypothetical representation is yielded by MDS:
According to Shepp and Swartz (1976), Smith and Kemler (1977 & 1978) and Shepp (1978), pairs (A & B₁) and (A & C₁) in Diagram 4 are "correlated" (see also §9.2), since the respective values along one of the dimensions are identical.

If the choice of B₂ with respect to A is the same as the choice of C₂ with respect to A, the two dimensions are equally attended to. Or simply if A, B₂ and C₂ are chosen the same number of times, the two dimensions are equally attended to. If the subject attended to one of the two dimensions asymmetrically, it would predict that the choice of C₂ is greater than the choice of B₂ and vice versa. On the other hand, the choice of C₂ and B₂ is identical with respect to A, when the two dimensions are separately attended to.

Similarly, pairs (A-B₁ & A-B₃) and (A-C₁ & A-B₃) stand for "orthogonal" pairs, according to Shepp, Swartz, Smith and Kemler. These pairs are defined as orthogonal, since the relationship between A and B₁ is not diagonal to dimension 1, but orthogonal. (Correlated pairs are related to each other diagonally to the relevant dimension.) Likewise, the relationship between A and B is orthogonal to dimension 1. Similarly, the relationship between A & C₁, and between A & C₃ is orthogonal to dimension 2, as opposed to the diagonal relationship between A and C₂.
with respect to dimension 2. Now, suppose that the interpoint distances between \( A \) & \( B_1 \), between \( A \) & \( B_3 \), between \( A \) & \( C_1 \) and between \( A \) & \( C_3 \) are all equal. Separable processing would predict that the choices of \( B_1 \), \( B_3 \), \( C_1 \) and \( C_3 \) would be the same, since the two dimensions are equally attended to. As the above hypothetical cases illustrate, the experimental demonstration of separable processing requires special pairings of correlated pairs or orthogonal and equi-distant pairs. Unless the experimental materials are artificially designed to meet these requirements, it is difficult to demonstrate the presence of separable processing. MDS involving natural materials, as in the present study, hardly ever produces the correlated and orthogonal equidistant pairs. In this pilot study, some items happen to show some resemblance to correlated pairs. Therefore, the separable processing is suggested below, but it remains speculative.

Some explanation concerning Thurstone's and Torgerson's method of measurement is supplied in pilot study (4) (chap. 7) and chap. 5. It is not repeated here. This pilot study makes use of the simplest method, i.e., condition IV; Torgerson, 1958 - see also chap.7.

§4.3 pilot experiment

§4.3 is divided into three. §4.3.1 discusses what materials were used in this pilot experiment. §4.3.2 presents the method of elicitation for the MDS input data, since the MDS requires a special kind of data (i.e., the ranking of all possible paired items). §4.3.3 explains the method of how the judgement of translatability was elicited.

§4.3.1 materials used

This experiment deals with four verbs: kowasu (break), kudaku (smash), oru (bend) and yaburu (tear). The choice of these four verbs is determined by Kellerman (1977, 1978 & 1979). Kellerman used 9 Dutch sentences involving break.
In order to translate these 9 sentences into Japanese, the five verbs are necessary. One of Kellerman's sentences "His voice broke" was omitted for the following reasons. The sentence translates into Japanese thus: koe-gawari (voice-change) ga (subject-marker) shi-ta (do/occur + past tense marker). Therefore it involves (1) a nominalized expression and (2) the general verb suru (do). Suru is semantically remote from the four verbs mentioned above. The nominalized gawari (i.e., [k] in kawari (change) is altered into [1] due to the vowel (e) which occurs before (k)) relates to 'a change of state' in such "instantaneous inchoative verbs", as "break". But nominalization is a complex issue and beyond the scope of the present study. For these two reasons (1) & (2), the sentence "His voice broke" was not considered here.

Japanese sentences which correspond to Kellerman's 8 sentences are circled in the lists (in 1.3, Appendices). Each list of NL sentences contains idiomatic usages in Japanese which appear to cover the 4 quartiles in the MDS result presented in Kellerman's study. That is, each of the four NL verbs has literal and non-literal uses as well as taking concrete and abstract nouns as subjects or objects. Some basic grammatical information and translations are supplied in appendix 1.

In preparing the experimental lists, the following two points are considered:

(1) Since the subjects are adult speakers of Japanese, it is expected that they are very sensitive to refined semantic differentiations. In order to make the NL sentences less ambiguous, adverbial clauses or phrases are added to supply some contextual information. The contexts conveyed by these adverbial clauses and phrases can narrow down the meaning of the verbs so that the message may be clear to the mature subjects. (In the main study these contextualizing clauses are omitted because of the danger of confusing the much younger subjects involved.)
The case relations and grammatical relations which occur in Kellerman's 9 sentences were studied (see appendix 2), in order to have some clear idea of what has to be learned by Japanese learners. This preliminary analysis made me wonder what the effect of mixing case relations of surface subjects might be. In Kellerman's study the 'cases' of the surface subjects are different, in addition to the use of the adjectival form: e.g., "Since the accident he's been a broken man". Since the latter would bias the translatability judgements, the adjectival form was not used here.

One inquiry was made among 23 EFL teachers to see whether or not the 'case' variation in Kellerman's study can affect people's comprehension regarding the meaning of the main verb 'broke'. The 4 sentences, each containing a different case as its surface subject were given to these EFL teachers:

a. She (agentive) broke his leg.
b. His leg (objective) broke.
c. The iron bar (locative/instrument) broke his leg.
d. The slip (abstract locative) broke his leg.

These sentences were paired (6 pairs in all) and the EFL teachers rated the similarity of meaning of "broke" on the 5-point scale: (1) identical; (2) close relationship; (3) some relationship; (4) slight relationship; (5) very different.

In all four sentences the objects which are affected by the action of 'breaking' are all either locatives or patients; the objective (nominative in Anderson's terminology) in sentence c can be 'instrument' in a special context where somebody deliberately broke his leg using the iron bar. Sentences a, c & d emphasize the circumstantial differences concerning what and who caused the breaking. The deliberate choice of surface subject on the part of a speaker is thus a useful means of conveying how much information about the agent responsible for the
action a speaker wishes to convey (see van Buren, 1975). If he does not intend to convey who is responsible, he would use sentence b. Sentence b conveys the minimum information about the bare fact of who did the action and what he broke. The meaning of 'break' in this sense does not require any circumstantial information. If so, the meaning of 'break' in all 6 pairs would be assessed by subjects as identical. But the understanding of meaning also depends whether the subjects regard "break" as intrinsically an intransitive or transitive verb. If they regard the verb as essentially transitive, their ratings would differ, depending on what is seen by the subject as a typical subject: cf. Dik points out the likelihood of surface subjects in terms of the hierarchy of seven case relations (1982). Thus, it is likely that the surface subjects whose case relations are varied, as in Kellerman's study, are the sources of different responses.

Only 4 out of 23 subjects rated all 6 pairs as (1), i.e., identical. In spite of being warned to pay attention to purely semantic elements, 23 EFL teachers seem to have regarded the difference in the implied situations as affecting the meaning of "broke". It suggests that the verb is regarded as essentially transitive by the majority of these EFL teachers.

The point is that by mixing the cases of surface subjects we implicitly include one extra irrelevant experimental variable derived from syntactic variations. In other words, if we present the following pair as in Kellerman,

The cup broke;
He broke his leg,
the subjects may think there must be some special reason for choosing "He broke his leg", instead of "His leg broke", or for choosing "The cup broke" instead of "She broke the cup". For this reason, in constructing the experimental sentences the case roles of surface subjects are kept the same in all sentences as far as possible. When this is
not possible, the list as a whole is balanced between transitive and intransitive Uses. (In the main study, all sentences involve transitive uses so that less sophisticated subjects should not be distracted from the main experimental task.)

§ 4.3.2 the method of elicitation of input data for MDS and for the Thurstone-Torgerson method

40 Japanese teachers of English who attended the summer course at Edinburgh in 1978 took part in this elicitation. The age of the subjects ranged from the late twenties to over fifty. They were presented with 9 or 10 Japanese sentences composed of 4 verbs: kowasu (break), kudaku (smash), oru (bend) and yaburu (tear). In total, they dealt with 37 sentences. The presentation of the four lists was counter-balanced: 10 subjects dealt the lists in the order of break, smash, bend and tear; 10 subjects, in the order of tear, break, smash and bend; 10 subjects, in the order of bend, tear, break and smash; and 10 subjects, in the order of smash, bend, tear and break. They were given the instruction sheet (see 1.0, Appendices) and some examples from English sentences were presented on the blackboard. They were told to rate all possible pairs, if they could.

(.....see Instruction sheet in 1.0, Appendices)

The task involved four procedures:

(1) grouping the NL sentences according to the similarity of meaning;

(2) rating adjacent pairs on the 5-point scale;

(3) rating the relationships between groups;

(4) rating all possible pairs.

The subjects spent 30 to 40 minutes on the task. No time-limit was specified. Some who did not complete the task within the experimental session returned their answer sheets after a week. In this sense, the elicitation was not
controlled.

On the basis of their rating, Torgerson's equation was used to compute the 'feel' of similarity between all possible pairs. The result of this measurement of similarity is abbreviated as SIM below. Since MDS requires as input the rankings of all possible combinations of paired items, SIMs were ranked and the ranked SIMs were used as input to MDS. The MDS used here is the same MINISSA program as in Kellerman (1977 & 1978).

§4.3.3 the method of elicitation of the translatability judgement

In order to make a plausible assessment of 'sure' items, three factors are considered:

(1) Whether the item has been specially taught at educational institutions.

(2) Whether it is the most frequently chosen item by most subjects.

(3) Whether it is consistently chosen by the same subject over several trials.

Since the third factor relates to the elicitation of the judgement data, it is discussed below. However, the first factor (1) needs to be briefly touched on. The 15 subjects (post-graduates) who took part in the judgement test had learnt English in Japan, where textbooks used at high schools and junior high schools are much more uniform than in other countries. The Japanese Ministry of Education issues specific guidelines on both the vocabulary and grammatical constructions to be taught at each level, as well as the number of hours to be spent on each point. Moreover, the textbooks are all authorized and examined by the Ministry of Education. Thus, we can get a fairly good picture of what is specially taught in Japan by looking at the textbooks. The major textbooks which were used in the areas the subjects came from are consulted and some examples are given in appendix 4.
The third factor (3) mentioned above requires us to specify how many trials are needed for us to claim some consistency of choice by the same subject. The number of trials can be determined by evaluating random events. There are 9 or 10 items in each list. In the case of the 9-item list, there are 512 possible combinations. The choice of a single item yields 256 possibilities. Thus, the probability of choosing any single item is merely 0.5, which is a random event. But if the same test is repeated and, the test & retest are "independent", the probability of choosing the same single item is 0.25, which is less random than the test consisting of one trial. If the same test is repeated once again and the tests are independent, the probability becomes 0.125. Thus, the more independent trials there are, the less the likelihood of pseudo-consistency.

There were 2 to 5 correct answers in the lists presented to the subjects. Supposing that the subjects have chosen the correct items as translatable. The choice of these correct answers might have been the outcome of either their legitimate 'formal' learning, or simply the result of 'NL transfer'. Or they could have acquired these well-formed TL expressions while they have lived in the English speaking country. But, we hope as EFL teachers that the 'correct' answer is the result of good teaching and learning. For this pedagogical reason, we wish to avoid any possibility of 'correct' choices being designated 'inconsistent' and therefore 'unsure' items. This suggests that there should be a criterion of consistency so that at least our operational definition of consistency enables us to make all correct items legitimate candidates for 'sure' items. Thus, it is reasonable to determine the number of trials by the number of correct answers there were in the lists to be presented to the subjects. The probability that 2 to 5 correct items can be chosen at random in each trial is estimated as follows (see, Feller, 1970):
<table>
<thead>
<tr>
<th></th>
<th>2 responses</th>
<th>3 responses</th>
<th>4 responses</th>
<th>5 responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>0.2500</td>
<td>0.1250</td>
<td>0.0625</td>
<td>0.0313</td>
</tr>
<tr>
<td>Trial 2</td>
<td>0.0625</td>
<td>0.0156</td>
<td>0.0039</td>
<td>0.0009</td>
</tr>
<tr>
<td>Trial 3</td>
<td>0.0039</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Trial 4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

As it is reasonable to set \( p = 0.01 \) as our rejection point, it was decided that the number of trials would be three.

The above level of consistency is one of the three factors which determine the 'sure' item in the present experiment. But this operational definition of 'consistency' has another implication. The present definition of consistency will differentiate the variable performance from the systematic performance. In this sense, the present operational definition relates to the usual notion of IL competence (see chap. 2 and Language Learning (1976) special issue). The systematic performance has been regarded as reflecting IL competence, while the variability and fluctuation characterises 'performance'. For this reason, the present definition is one possible way of finding out this distinction between performance and competence. So, we can summarise the above criterion of consistency as having three features:

1. The 'sure' item will always be counted as 'consistent'.
2. The probability of occurrence of variable performance counted as 'consistent' is sufficiently low for the present purpose.
3. The consistency criterion may be related to the distinction between the IL competence and performance.

15 Japanese speakers who had lived in Britain more than a year took part in this translatability test. They were presented with the four lists consisting of 37 Japanese sentences and asked whether each Japanese sentence could be rendered in English using 'break'. They were then asked to complete each list in 2 minutes. The test was repeated three times on the same day, the order of presentation being different for each trial. Two
distractive tasks were introduced between the two consecutive trials: the subjects had to answer questions about the verbs "be" and "have". The distractive tasks took them longer than the translatability test (about 20 minutes).

§4.4 result & discussion:

As indicated in §4.0, the translatability data was analyzed in three ways: I (in §4.4.1), II (in §4.4.2) & III (in §4.4.3). The MDS produced the configurations listed in 1.3, Appendices. The information necessary for the analysis I, II & III are presented in 1.1 - 1.3 Appendices. The frequency tables summarizing the consistent and variable choices made in the translatability test are given in 1.1, Appendices.

§4.4.1 Analysis I

This analysis takes up Kellerman's data treatment. As in Kellerman, we regard the MDS results as presenting an adequate representation of the NL interlexical structure. But, as we have seen in §4.0, it is implausible to suppose that the NL interlexical structure is transparently reflected in the judgement data. Rather, the judgement of translatability should be seen as reflecting the interaction between the specific form of processing the subject adopts and a 'sure' item whose translatability the subject is confident of. But, Analysis I replicates Kellerman's data treatment and excludes the latter effect of 'sure' item from our consideration. Thus, in this analysis, we consider the former effect of processing upon the judgement data. This is a modification of Kellerman's data treatment, since his analysis does not involve the effect of processing. But analysis I is the closest to his method of the three.

As we have seen above, Kellerman considers the rank-order correlations of his translatability data with one of the two or three dimensions in the MDS results he obtained.
This data treatment suggests that Kellerman implicitly assumes an asymmetric separable processing. This is because when the subject adopts this processing, he favours only one dimension, by definition (see §4.2). In accordance with Kellerman's assumption, analysis 1 assumes also an asymmetric separable processing. We are not concerned here with the question of which is seen as the 'sure' item by the subjects.

Table 2 presents Spearman's rank order correlations between the respective values along dimensions 1 & 2 and the judgement data. The latter data are divided into two types of response: consistent and variable (see §4.3.3 in which a motivation for this classification of responses is presented). The term 'consistent response' refers to the cases in which the same subject judges the same item to be translatable over the three trials. 'Variable responses' are those which did not satisfy the definition of consistency.

<table>
<thead>
<tr>
<th>types of verb</th>
<th>dimensions</th>
<th>consistent</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>kowasu (break)</td>
<td>DIM1</td>
<td>p=0.625</td>
<td>p=0.7583</td>
</tr>
<tr>
<td></td>
<td>DIM2</td>
<td>p=0.3167 (N.S)</td>
<td>p=-0.3167 (N.S)</td>
</tr>
<tr>
<td>kudaku (smash)</td>
<td>DIM1</td>
<td>p=0.9033</td>
<td>p=0.9833</td>
</tr>
<tr>
<td></td>
<td>DIM2</td>
<td>p=-0.025 (N.S)</td>
<td>p=0.025 (N.S)</td>
</tr>
<tr>
<td>oru (bend)</td>
<td>DIM1</td>
<td>p=-0.0088 (N.S)</td>
<td>p=-0.1604 (N.S)</td>
</tr>
<tr>
<td></td>
<td>DIM2</td>
<td>p=0.4923 (N.S)</td>
<td>p=0.5582 (N.S)</td>
</tr>
<tr>
<td>yaburu (tear)</td>
<td>DIM1</td>
<td>p=-0.225 (N.S)</td>
<td>p=0.1 (N.S)</td>
</tr>
<tr>
<td></td>
<td>DIM2</td>
<td>p=0.0833 (N.S)</td>
<td>p=0.4333 (N.S)</td>
</tr>
</tbody>
</table>

Table 2
(p stands for ρ.)

111
Table 2 shows that in the case of kowasu and kudaku, the judgment data correlates significantly better with DIM1 than with DIM2. We regard the translatability judgement as the outcome of a specific type of processing the subjects adopt. Analysis I assumes the asymmetric separable processing. The fact that DIM1 yielded the superior correlations with the judgement data may indicate that this dimension was favoured by the subjects; as we have seen previously, in the asymmetric separable processing there is by definition one compelling dimension to which the subject's attention is drawn. If so, it may follow that the asymmetric separable assumption of Kellerman's may be justified in these data (kowasu & yaburu). However, the judgement data for oru and yaburu do not support this interpretation. As Table 2 indicates, neither of the two dimensions yielded significant results, suggesting that no dimensions were specially favoured by the subjects. While kowasu and kudaku justifies Kellerman's implicit assumption, the other two verbs do not replicate this justification.

Kellerman argues that a learner develops the 'feel' of translatability based upon his intuition of 'core' items. In order to substantiate this claim, he draws a reader's attention to the fact that one of the MDS dimensions can be called 'core/non-core dimension'. His data shows high correlation with this dimension. In this sense, this dimension serves as a crucial notion in his argument. For this reason, it is important for us to examine whether one of the MDS dimensions in this pilot study resembles Kellerman's core/non-core dimension. When we look at the configurations for kowasu depicted by the MDS, it appears to be reasonable to regard this DIM1 as more similar to Kellerman's core dimension than DIM2, since DIM2 clearly indicates a concrete-abstract dimension:

---

*Fig. 1 in 1.0 & Fig. 1 in 1.2, Appendix*
However, there is one difficulty in regarding DIM1 as the core dimension. Items 8, 4, 1, 2 & 6 run almost in parallel with DIM2. They do not show any obvious differences in terms of the "coreness" values (see the respective values along DIM1 & DIM2 in Fig. 1, 1.3, Appendices). Because the values along DIM 1 do not differ much, these items almost fall along a line in parallel with DIM2. This indicates that there is no typical core item which makes our interpretation, or rather our naming, of the dimension objectively convincing, (although the user of MDS is, as indicated above, encouraged to name dimensions by looking at configurations.) In the MDS result of kowasu, all 5 items (8,4,1,2 & 6) may be called 'core' equally well. This is counter-intuitive, since items 1, 2 & 6 are metaphorical and are far from our usual sense of 'core'. However, there is some ground in saying that kowasu supports Kellerman's claim about his 'core' assumption, since, in comparison with DIM2, DIM1 is at least more like his core dimension.

As we have seen above, kudaku supports Kellerman's
asymmetric separable assumption. For this reason, we will examine whether Kellerman's assumption of 'core' holds true in the case of kudaku. The MDS configuration for this verb is given in 1.3, Appendices.

For this reason, we will examine whether Kellerman's assumption of 'core' holds true in the case of kudaku. The MDS configuration for this verb is given in 1.3, Appendices.

<table>
<thead>
<tr>
<th>DIM2 (concrete)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 wind</td>
</tr>
<tr>
<td>6 waves</td>
</tr>
<tr>
<td>5 enthusiasm</td>
</tr>
<tr>
<td>8 philosophical concept</td>
</tr>
<tr>
<td>4 cup</td>
</tr>
<tr>
<td>7 atmosphere</td>
</tr>
<tr>
<td>1 attitude</td>
</tr>
<tr>
<td>(abstract)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 spirit</td>
</tr>
<tr>
<td>9 mind</td>
</tr>
<tr>
<td>(non-core)</td>
</tr>
</tbody>
</table>

Fig.2 kudaku

(.....see also Fig. 3, in 1.3, Appendices)

This verb conforms to Kellerman's interpretation of the 'core' better than kowasu does. Item 4 (cup) is a good example of the literal use of this verb, and is placed at one extreme of DIM1. Further, four items 5, 8, 7 & 1 which illustrate the metaphorical uses of this verb are positioned at the other extreme of DIM1. In this sense, DIM1 may be named 'core/non-core dimension'. The presence of one typical exemplar for the basic meaning (i.e., item 4) at one extreme of DIM1 certainly makes our naming more plausible in kudaku than in kowasu.

But, there is one problem in regarding DIM2 as 'concrete-abstract'. Items 5 & 8 (enthusiasm & philosophical concept) are clearly abstract, but they would be included among 'concrete' items, if we regarded DIM2 as the concrete/abstract dimension. This illustrates a general
difficulty in interpreting dimensions. The same point has been made by the originator of non-metric MDS (Shepard, 1974 and Carroll & Shepard, 1976). However, since DIM1 could be associated with Kellerman's core dimension, we can regard kudaku as partially supporting Kellerman's hypothesis.

So far we have seen that one dimension correlates better with the judgement data in kowasu and kudaku, which justifies Kellerman's implicit assumption of asymmetric separable processing. We have also seen that Kellerman's 'core' assumption may be partially justified in these verbs. However, the other two verbs (oru and yaburu) do not support Kellerman's implicit assumption, since neither of the dimensions show significant correlations with the judgement data. These disconfirming results can be interpreted two ways. First, the assumption of asymmetric separable processing might have been too simple to account for the judgement data. Although the asymmetric separable processing assumes that there is one compelling dimension in operation, the other dimensions could be contributing to the judgemental process the subjects carry out (see analysis II & III). Second, the negative results in oru and yaburu might be due to the nature of these verbs. This point is examined below.

First, the four verbs are briefly compared in terms of the basic meanings. Table 1 in appendix 3 presents an informal account of 13 verbs which are related to the four verbs. It is intended to show that Japanese has a systematic means of lexicalization, depending upon the inherent features of objects, motor movements required and the resultant state which the activity brings about. Only in the context of these related 13 verbs, can one understand the distinctive features of the four verbs in their literal use. These four verbs are circled in the table. The first column of the table describes the characteristic nature of the objects associated with each verb. The second column (Action) analyzes the action. The third column (Result) shows how the result of the action relates
to the lexicalization. The second column is subdivided into four parts: (a) the direction of force; (b) the main motor movements involved (These relate to the shape, size and quality of an object); (c) the scope describes whether the force extends over the whole or part of the object; (d) the instrument specifies tools which an average Japanese tends to associate with the activity, when the instrument is not overtly expressed.

As Table 1 in appendix 3 shows, the differential lexicalization of an intended message in Japanese largely depends on the nature of the object and the effect of the activity. Kowasu typically occurs with artificial objects such as cars, furniture, boxes and so on, that is, the objects which have functions in daily life. The result of the kowasu activity brings about malfunction and the shape is often deformed. This verb does not normally co-occur with natural objects, such as fruits, vegetables or plants.

Kudaku is typically used with objects which are brittle, such as a lump of sugar. This activity brings about the disintegration of the object. In comparison with the near-synonymous kuzusu (pull down), tsubusu (crush) and kaku (chip), the area in which the separation occurs is virtually unlimited. However, whether or not it covers the whole area depends upon the specific context and how far the speaker is conscious of the other near-synonyms at the time of utterance.

Oru (bend) is a more general verb than mageru (bend). The latter is used with the objects which are pliable and the result of the activity does not bring about the separation of the object. Oru is ambiguous in this respect; the activity may or may not result in the separation of the object. This is because the verb is used in three ways: (1) bend; (2) snap; (3) fold. However, these meanings to some extent form a conceptual unity. Folding a sheet of paper is the more precise form of bending it, i.e., bending double. On the other hand, in bending a stick, if the force we apply is more than it can sustain,
it snaps. These literal uses depend upon the nature of the materials and the force applied. Thus, the context usually clears up the meaning intended. For example, if the material is a sheet of paper, the expression of oru stands for folding rather than snapping it. Yaburu (tear) is usually used with objects which are sheet-like such as paper & cloth. The usage is very similar to the English verb "tear". The related verbs are saku (split), waru (crack), and chigiru (tear off). In saku the action of tearing typically extends over the whole area, while in chigiru a torn part tends to be a small portion of the whole. Yaburu does not involve this differentiation. While waru is used for non-sheet-like objects such as eggs and rice crackers, yaburu occurs with a sheet-like object.

So far, we have seen the literal use of the four verbs. The non-literal co-occurrent uses are discussed below. The discussion is based upon the idea of "co-occurrence restriction". This means here that we try to find some notional categories which can account for what kinds of noun phrases typically co-occur with the four verbs. Then, we try to see whether these notional categories can account for the groups which appear in the MDS results. If these notional categories are similar in all four verbs, it would mean that the discrepancy we have noted above (ie., the better correlations with the judgement data in kowasu and kudaku, as opposed to the poor correlations in oru and yaburu) might not have had much to do with the nature of the verbs.

Following a few simple notional criteria, the noun phrases are purely mechanically grouped. This is because, once we introduce various contextually governed criteria (such as the primary purposes of objects in daily use in order to show how they are used metaphorically,) any object tends to show overlapping purposes and even the primary function is not easily defined. In order to avoid unnecessary details, the following self-explanatory criteria are used.

A. body-part terms
B. perceptible event or state: e.g., silence
C. metaphorical use derived from body-part terms: e.g., heart, nose
D. psychophysical event: e.g., conversation
E. metaphorical use derived from a psychophysical event: e.g., dream (used in the sense of a wishful plan).
F. abstract:
   i (+ subjective): e.g., plan (in the sense of a personal schedule)
   ii (+ interpersonal): e.g., promise
   iii (+ institutional): e.g., rule, law
   iv others: there is only one example of this in the experimental list, i.e., "difficult philosophy".

The non-literal uses in kowasu are classified notionally as follows. The literal translations of the noun phrases are given first. When the literal translations were not used in translating the complete sentences, the actual phrasings used in translation are presented in brackets. The numbers heading the noun phrases are those used in the experimental lists which are given at the top of the respective MDS results in 1.3, Appendices.

kowasu

A: (3) abdomen (stomach)
   (5) body (health)
D: (7) mood (my brother)
E: (9) dream
F: (i) (1) plan
   (ii) (2) arranged marriage
   (6) negotiations

When we compare the literal use of kowasu with the non-literal use, it can be said that the latter tends to extract some of the basic features and to emphasize these aspects: (this may be related to "analogy", "shift" and "transfer"
in the development of word meaning. cf Waldron, 1979, chapters 6-9; Kunihiro, 1982, chapter 3.) For instance, kowasu with body-part nouns does not require the selectional feature (+ artefact) or those listed in (+ Action) of Table 1 in appendix 3. It only emphasizes a part of the result, i.e., the state of malfunction. Likewise, kowasu with abstract nouns relates only to a change of state. It still carries a basic semantic feature of (+ destruction), conveying not a neutral change, but a negative effect which the change brings about. Including the literal uses of sentences (4) & (8), the above four categories can account for the clusters presented by the MDS result.

Similarly, the sentences in the kudaku list are grouped as follows:

B: (3) the force of the wind
(6) the waves used in the sense of the impact of waves
D: (7) the atmosphere of the gathering (breaking ice)
(1) his attitude used as a third person's observation about the man in question (his corners)
E: (2) heart a used in the sense of suffering (breaking his spirit)
(9) heart b used in the sense of deep commitment
taxing my brain for the company's benefit;
 cf some similarity with the English expression "putting my heart into it")
F: (i) (5) enthusiasm
(iv) (8) the difficult philosophy (difficult philosophical concepts)

The notional categories fit in with the five clusters in the MDS result, once we include the literal sentence of (4).

![Graph](image)

**Fig. 4 kudaku**

(.....see also Fig. 3 in 1.3, Appendices)

It will be recalled that we are assuming that the MDS result offers an adequate account of the NL interlexical structure. At present we are discussing how we can describe the groups in the learner-generated map. If the notional classification mentioned above can account for the other two verbs oru and yaburu equally well, it would mean that as far as the MDS results go, oru and yaburu which yielded no significant correlations with the judgement data are

120
similar to kowasu & kudaku which yielded significant correlations.

Yaburu sentences are grouped as follows and they correspond to the clusters in the MDS result.

B: (6) silence
C: (2) heart
E: (7) dream
F: (i) (9) plan
   (ii) (8) promise
   (iii) (1) rule
      (4) world record

abstract

<table>
<thead>
<tr>
<th>8</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

a body-part

<table>
<thead>
<tr>
<th>metaphor</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>(7)</td>
</tr>
</tbody>
</table>

psychophysical

<table>
<thead>
<tr>
<th>metaphor</th>
</tr>
</thead>
<tbody>
<tr>
<td>(6)</td>
</tr>
</tbody>
</table>

perceptible

Fig. 5 yaburu

(.....see also Fig. 2 in 1.3, Appendices)

The sentences with oru involve the following four categories.

A: (4) leg
   (7) arm

B: (8) road used in the sense of direction

C: (1) bone - idiom (pains)
   (2) ego used in the sense of self-assertion (giving in)
   (6) nose - idiom (pride)

D: (9) conversation
The MDS result for oru presents an unexpected relationship between the non-literal items (7 & 4) and the literal items (10, 3 & 5). According to the notional categories, the body terms (7 & 4) are regarded as non-literal. However, these non-literal items (7 & 4) are very close to the literal item (10), while the literal items (3 & 5) are further apart from the literal item (10): see Fig. 6. This peculiarity is not seen among the other three verbs. This peculiarity relates to three points. (1) The non-literal category of body-part terms may not be harmonious with the nature of the terms (7 & 4) used with oru. (2) Oru illustrates again the usual difficulty in interpreting the other dimension as core. This difficulty is already indicated by the case of kowasu. (3) In relating the MDS result with the judgement data, the clusters in the MDS configurations are as important as the dimensions. This point relates to the following finding that, although items 3 & 5 represent literal (and therefore basic) uses, no subjects regard them as translatable.

Since we are currently examining whether the results in oru and yaburi, which did not support Kellerman's hypothesis are due to the nature of these verbs, point (1) is dealt with first. As we have seen, oru involves three uses which are
conceptually similar to three English words: (1) bend, (2) snap and (3) fold. Depending on the nature of the object and the force applied, the three senses are differentiated. In this sense, the context of situation which specifies what is the object and how the force is applied in a given situation facilitates this differentiation. Sentence (7) is provided with a context, while sentence (10) is not (see below). Thus, oru in (7) indicates bending, but (10) is ambiguous.

(7) ude wo ottari nobashi-tari
    (arm) object marker oru+iteration stretch+iteration
    shi-ta.
    do + past tense
    (I bent and stretched my arm over and over again): see appendix 1.
(10) kasa no e ga ore-ta
    umbrella of stem subject-marker oru+past tense
    (The stem of an umbrella was bent): see appendix 1.

In (7), although arm has a joint whose natural purpose involves bending, without the contextual information of "stretching the arm", the simple combination of ude and oru becomes ambiguous. "Ude wo oru" can mean either bending or breaking the bone of the arm. Sentence (10) is similar to this simple combination and no extra information is supplied. Thus, oru in (10) can mean either "bent" or "snapped". By reading sentence (10), we normally think that it would be the case of a stem being bent, since snapping requires unusual force.

In the following example, sentence 1 indicates that the stick is snapped or bent. Sentence 2 indicates that, since an elbow has a joint which is intended to be bent, the verb stands for bending rather than snapping. On the other hand, in sentence 3, although it has a joint, the leg's primary purpose is not to be bent. Thus, our unmarked reading tends to suggest the case of the leg bone being snapped (i.e., broken).

1. bō wo oru.

123
It may be noticed that, when body-part terms occur with oru, they are not used in a completely literal way. Because it is not the leg but the bone of the leg that is broken or bent, is an example of "metonymy". (e.g., "The kettle is boiling" means the water in the kettle is boiling: see Waldron, 1979: 186). In this sense, it is not a literal use but a figure of speech. This is the reason why we have included body-part terms as belonging to the non-literal category.

However, when we compare the body-part terms used with oru with those used with kowasu, we recognize the difference. The former is more literal than the latter. Since the literal use with kowasu requires the feature (+ artefact) and body-parts are not artificial objects, they are more legitimately regarded as non-literal. On the other hand, the literal use of oru in the case of snapping tends to be associated with the feature (+ stick) (e.g., the branch of a tree, pencil, chopsticks), but since oru is used for "folding" as well, it is more difficult to specify the distinctive feature which is inclusive of bending, folding and snapping (Kunihiro, 1970, 1982; Morita, 1977). However, the learner-generated MDS map shows the closeness of body-parts terms (7 & 4) to the literal use of item (10):

![Diagram](image-url)

Fig.7 oru

124
Since we are concerned with the learner-language, it is important to appreciate that items 7 & 4 are clustered close to the literal use (10).

Table 3 below summarizes the foregoing discussion concerning the nature of the four verbs. The discussion was restricted to the question of whether the four verbs co-occur in notionally dissimilar or similar categories. We have dealt with the six categories. Our intention was to see whether these categories are adequate to explain the clusters produced by the MDS. The six categories can be applied equally well to kowasu, kudaku & yaburu and conditionally to oru. The discussion shows that these categories account for the clusters in the MDS results.

<table>
<thead>
<tr>
<th>literal</th>
<th>non-literal</th>
</tr>
</thead>
<tbody>
<tr>
<td>verbs</td>
<td>A</td>
</tr>
<tr>
<td>kowasu</td>
<td>4,8</td>
</tr>
<tr>
<td>kudaku</td>
<td>4</td>
</tr>
<tr>
<td>oru</td>
<td>3,5,10</td>
</tr>
<tr>
<td>yaburu</td>
<td>3,5</td>
</tr>
</tbody>
</table>

A: body part terms  
B: perceptible  
C: body part metaphors  
D: psychophysical  
E: psychophysical metaphor  
F: Abstract


Table 3
Since we are concerned with the interlexical NL structure, it is a meaningful test to examine the MDS results in terms of the categories derived from the method of co-occurrence restriction. There are empty cells in Table 3. In some cases, it is possible to fill them in, but it is not possible to do so in all cases - unless we are prepared to accept highly metaphorical unusual uses. However, the better correlation with the judgement data observed in kowasu and kudaku, in contrast with the poor correlation in oru and yaburu, does not appear to be due to the specific nature of these verbs. Nor does the poor correlation with the judgement data in oru and yaburu appear to be related to the specific nature of these verbs. It is at least clear that yaburu does not support Kellerman's core hypothesis.

As indicated above, not only kowasu but also oru pose some difficulties in specifying which of the two dimensions is a "core" dimension.

\[
\begin{array}{c|c|c|c}
\text{DIM2} & \text{DIM1} & \text{DIM2} \\
\hline
\text{literal} & \text{body parts} & \text{literal} \\
8 & 4 & 10 \\
3 & 2 & 7 \\
\text{abstract} & 5 & 6 \\
1 & 8 & 9 \\
\end{array}
\]

Fig. 8 kowasu

Fig. 9 oru

In kowasu, if we regard DIM2 as the core dimension, DIM1 becomes either concrete/abstract or abstract/concrete. The former interpretation entails that the abstract items of 1, 2 & 6 would be included in the 'concrete' cluster. This is obviously incorrect. In the latter interpretation, the literal and concrete group of 4 & 8 would stand for 'concrete items'. For this reason, we have regarded DIM1 as the core dimension. Then we have several core items,
since 4, 8, 1, 2 & 6 do not differ much along dimension 1. Items 1, 2 & 6 are metaphorical and they are not 'core' items in any sense. Oru also poses the similar problem to the case of kowasu above. In oru, items 10, 3 & 5 are literal items. Both dimensions relate to these three items. It is difficult to determine which is the 'core' dimension. Even when we regard the body-part terms of 7 & 4 as semi-literal, we are still facing the same difficulty of interpreting the core dimension. Thus, both kowasu and oru appear to challenge Kellerman's hypothesis which attaches great importance to this single dimension.

When we look at the judgement data (Tables 1-4, in 1.1, Appendices), it can be observed that some items are not chosen by any subject: for instance, items (7 & 8) in kudaku and items (1, 3, 5 & 8) in oru. The most interesting case is items (3 & 5) in oru. They illustrate a folding concept, one of the basic meanings of oru. In this sense they are possible candidates for core items and according to Kellerman's hypothesis, should be judged as translatable by the subjects. However, neither consistent nor variable responses show that they are regarded as translatable. This could be interpreted as a counter example to the core hypothesis. Further, this also indicates the importance of clusters in relating the MDS result to the judgement data. Since items 3 & 5 form one category derived from a folding concept, we could assume that the subjects treated items 3 & 5 as one group and judged the two together as untranslatable. Kellerman's own data also support this hypothesis of categorisation. In Kellerman (1977) the MDS result shows two distinct clusters. When the correlation with the judgement data was computed separately according to the respective cluster, the rank order correlations improved surprisingly: see §4.1 where it is indicated that his original data treatment yielded the lower correlation when the clusters were not considered. Although Kellerman has not pursued the important role of clusters in latter papers, it is important to bear in mind a possibility that, presented with a list of sentences, the subjects might categorize the items in their own way. If the items in
the list are arranged randomly, it is likely that they would group the items according to their conceptual categories while judging the translatability. If this happens, the clusters represented by the MDS result are more important than the dimensions themselves. This possibility of categorization is examined in pilot study (3). The present pilot study also discusses the topic below.

§ 4.4.2 Analysis II

This analysis examines whether Kellerman’s hypothesis of coreness can hold under the assumption of integral and asymmetric separable processings, in combination with the assumption of 'sure' item. We have dealt with Kellerman's implicit assumption of asymmetric separable processing in Analysis I. But, when we introduce the notion of 'sure' item, the same assumption can be considered differently. Taking the following hypothetical space, the difference is explained first.

```
<table>
<thead>
<tr>
<th>DIM1</th>
<th>DIM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>-1</td>
</tr>
<tr>
<td>D</td>
<td>-2</td>
</tr>
</tbody>
</table>
```

Let us suppose that A is a "sure" item. The asymmetric separable assumption would indicate that one of the two dimensions yields a better correlation with judgement data than the other dimension does. We assume as before that the analogical process operates in the judgemental process. So we first measure the interpoint distances between A and the rest of the items along DIM1 and DIM2 separately. Then,
we rank the interpoint distances. Second, we rank the judgement data. Third, we compute the rank-order correlations with the ranked judgement data and examine which dimension yields a better correlation.

<table>
<thead>
<tr>
<th>interpoint distances</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/B</td>
<td>1</td>
</tr>
<tr>
<td>A/C</td>
<td>2</td>
</tr>
<tr>
<td>A/D</td>
<td>3</td>
</tr>
</tbody>
</table>

This procedure is different from Kellerman's data treatment, although the hypothetical example happens to accord with his core hypothesis in that if DIM2 is a core dimension, the hypothesis regards the order of B, A, C and D as predicting the subject's judgement of translatability. Since Kellerman's study did not consider integral processing at all, it is interesting to see what result we can obtain. Once the 'sure' item in each verb is specified, the assumption of integral processing leads to a straightforward data treatment. As we have seen in §4.2, MDS computes all interpoint distances between paired items. One can make use of this information and calculate the rank order correlations of the interpoint distances between the 'sure' item and the rest of the items with the ranked judgement data. As mentioned previously, analysis II assumes that the MDS result gives an adequate representation of the NL interlexical structure.

The operational definition of 'sure' item is given in §4.3.3. It will be remembered that if all three of the following criteria are met, the item is regarded as 'sure'. (1) the item is specifically taught in educational institutions (see, appendix 4); (2) the item is consistently chosen over the three trials by the same subject; (3) the item is most frequently chosen by the majority of subjects. For the third criteria the respective sum of consistent and variable responses is considered separately and a chi-square test is used to see whether there is any significant difference. The 'sure' item in the four verbs is determined according to these criteria.

Result

129
The items in kowasu regarded as "sure" are 4(cup) & 8(umbrella). Although the specific combinations of "break" with "cup" and "break" with "umbrella" did not appear in the textbooks, the first examples of "break" are in 3rd year junior high school textbooks and occur with domestic utensils; e.g., vase (see appendix 2). Items 4 & 8 are judged as translatable by the majority of the subjects and are the most frequently chosen (a chi-square test is N.S). There is no reason to regard "breaking the cup" as surer than "breaking the umbrella". Thus, the interpoint distances are averaged over items 4 & 8. The rank order correlations with the judgement data are as follows:

kowasu (sure items: 4 & 8)

<table>
<thead>
<tr>
<th></th>
<th>Asymmetric separable</th>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>p=0.875</td>
<td>p=0.8214</td>
<td></td>
</tr>
<tr>
<td>DIM2</td>
<td>p=-0.4286</td>
<td>p=-0.4643</td>
<td></td>
</tr>
</tbody>
</table>

Integral

<table>
<thead>
<tr>
<th></th>
<th>MDS-SIM</th>
<th>p=0.3393</th>
<th>p=0.3036</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(N.S)</td>
<td>(N.S)</td>
</tr>
</tbody>
</table>

(p stands for \( \rho \).)

Table 4

As we have seen in Analysis I, DIM1 stands for the core/non-core dimension, but this naming entails multiple core items and in this sense Kellerman's core hypothesis is not fully supported. However, under the assumption of asymmetrically separable processing combined with the idea of 'sure' items, the superiority of DIM1 is observed, as in Analysis I.

It may be noted that the DIM1 result shows a positive correlation, while DIM2 shows a negative correlation. The latter is not statistically significant, but it is undeniable that DIM2 correlates negatively with the judgement data. The positive correlation of DIM1 proves that the degree of semantic closeness between the 'sure' items and a given item reflects the degree of acceptance as translatable. The more similar the items along DIM1 are, the more likely is the
given item to be judged as translatable. On the other hand, negative correlation with DIM2 means that the greater the semantic distance between the 'sure' items and a given item is, the more likely it is to be judged as translatable. (This is simply a paraphrase of the negative rank order correlation, while the more plausible interpretation is given below.) This seems to suggest that the subjects paid attention not only to DIM1 but also to some extent to DIM2 (since the negative correlation with DIM2 is N.S. and the semantic effect is reversed in DIM2). It will be recalled that separable processing represents the subject's ability to pay attention to each of the relevant dimensions equally well. In our situation, DIM1 is more dominant than DIM2, but there is no doubt that, DIM2 is contributing to the judgemental process in terms of the remoteness of DIM2 meaning. There is indeed some indication that DIM1 and DIM2 are separately attended to by the subjects. As the MDS result shows, items 1 & 6 and items 5 & 7 are correlated (see §4.2).

<table>
<thead>
<tr>
<th>Items</th>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlated(1 6)</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>correlated(5 7)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5

It should be noted that correlated pairs have a tendency to be judged as translatable by a similar number of people. Everytime the MDS results yield correlated relationships, they tend to show the same tendency (see below).

In the above we have seen that DIM1 correlates positively with the judgement data, while DIM2 correlates negatively. The interpretation of this was that the former indicates the role of similarity in the judgemental process and the latter, the role of dissimilarity. In other words, DIM1 positively contributes to the process of analogy and DIM2 negatively interacts with the process of analogy. This point can be related to the notions of assimilation & accommodation and of nativization and denativization (see chapter 131.
2.) NL Transfer is a form of assimilating TL to NL and is a type of nativization. On the other hand, by accommodating TL, denativization is achieved. The fact that DIM1 positively contributes to the judgemental process may be seen as a case of assimilation; one component of the NL structure is seen by the subjects as assimilatable. Whereas, the negative effect of DIM2 upon the translatability judgement may be likened to the accommodation process. That is, in order to accommodate TL, DIM2 which is a part of NL shows up negatively. So, the whole picture of the judgemental process can be seen in terms of the joint process of assimilating DIM1 (which is a part of NL) and accommodating TL by negating the analogical process in the order of similarity (which is reflected in the negative correlation of DIM2).

The present argument can be diagrammed as follows:

```
   * * *
    \   /
     \ / 
      v  
    assimilation

          DIM2

          \  
           \ 
            \ 
            DIM1

   \  \   \  \  
   \ |   |   |  
   | v   v   v |
   | accommodation |
```

(The 'sure' items are starred.)

Fig.11 kowasu

The above phenomenon involving assimilation and accommodation is more distinctly observed in kudaku.

The item in kudaku regarded as 'sure' is item 9 ("The cup was broken into pieces"). It is not only the case that the phrase "breaking something into pieces" is normally taught towards the end of 3rd year junior high school, but also
that it is most frequently chosen by the majority of subjects and that it is most consistently chosen by the same subject. Item 9 therefore meets the definition of "sure" item mentioned above.

\[
\begin{array}{ccc}
\text{kudaku (sure item : 4)} & \text{consistent} & \text{variable} \\
\text{DIM1} & p=0.8691 & p=0.9648 \\
\text{DIM2} & p=-0.7262 & p=-0.7109 \\
\end{array}
\]

Integrals
\[
\begin{array}{ccc}
\text{MDS-SIM} & p=0.8333 & p=0.9643 \\
\end{array}
\]

As in analysis I, kudaku shows that DIM1 correlates better with the judgement data than DIM2 does. But the assumption of integral processing is supported as strongly as the assumption of asymmetric separable processing. The point has a significant bearing on Kellerman's hypothesis, since integral processing claims that judgement is not the result of one dimension functioning as a criterial feature; rather, it is the interactive result of two or more dimensions.

As in kowasu, DIM1 correlates positively with the judgement data and DIM2 correlates negatively. Since the negative rank order correlation is obtained on the basis of the semantic closeness between the 'sure' item and the rest of the list, the negative correlation can be converted into a positive correlation on the basis of semantic remoteness. The more dissimilar to the 'sure' item a given item is (according to the respective values along DIM2), the more likely it is to be judged as translatable by the subject (and the effect is statistically significant). According to the result of DIM1, the more similar to the 'sure' item a given item is, the more likely it is to be judged as translatable by the subject. In this sense kudaku supports the joint contribution of semantic closeness (DIM1) and semantic remoteness (DIM2). As we have seen, it is a process of assimilating NL and accommodating TL. The former accepts NL similarity
and the latter rejects NL in order to accommodate TL. In this context it may be recalled that the rank order correlation under the assumption of integral processing is as good as those under the assumption of asymmetric separable processing. It is possible to connect this with the joint operation of assimilation reflected in DIM1 and accommodation reflected in DIM2. Since the correlations in both DIM1 and DIM2 are statistically significant and they jointly contribute to the judgemental process, it would naturally result in a high correlation under the assumption of integral processing. This may explain the high correlation obtained.

According to the MDS result, kudaku yielded more correlated pairs than kowasu (5 as opposed to 2). In addition, there were two orthogonal pairs (see §4.2).

<table>
<thead>
<tr>
<th>correlated pairs</th>
<th>consistent</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>{2, 7}</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>{7}</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>{7, 8}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>{5}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>{3, 9}</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>{2, 6}</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>orthogonal pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>{1, 7}</td>
</tr>
<tr>
<td>{7}</td>
</tr>
<tr>
<td>{7, 8}</td>
</tr>
<tr>
<td>{8, 5}</td>
</tr>
<tr>
<td>{8, 7}</td>
</tr>
</tbody>
</table>

Table 7

Apart from one correlated pair (2, 6), the 4 correlated pairs and the two orthogonal pairs support the prediction made
by assuming of separable processing. It will be recalled that separable processing predicts the similar number of choices among correlated and orthogonal pairs (see §4.2). Table 7 indicates the possible occurrence of separable processing, since the number of responses is either identical or highly similar. (In §4.2, for the sake of clarification, integral processing is described as the global and unanalytical mode, as opposed to the analytical mode of separable processing. But, as chapter 9 presents, integral processing is compatible with the analytical mode and therefore with separable processing. We have seen in this Analysis II that the integral assumption and the separable assumption are equally compatible with the data. It suggests that the subjects might have adopted the two processing at the same time. Since integral processing is compatible with separable processing (see §9.3.2(4), §10.1.4 and §10.2.2), the foregoing discussion of data is not contradictory.)

Kudaku also illustrates the importance of clusters in relating the judgement data to the MDS result. As we have seen in Analysis I, the MDS configuration shows some clusters:

<table>
<thead>
<tr>
<th>Items</th>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig.12

135
The subjects did not choose the two items belonging to the psychophysical category or the two items belonging to the abstract category (see Fig. 12). In making their judgements the subjects might have categorized the items and automatically rejected any category which they intuited as semantically remote from the 'sure' item. Particularly in the case of consistent responses, the frequency is zero (see Table on the right side of Fig. 12). This implies that the subjects reject a semantically remote category altogether as a group.

It is important to note that identifying the 'sure' item is the only possible way of assessing integral processing. Without this clue of the 'sure' item, it is not possible to choose one out of all possible combinations, since MDS simply computes all these possible interpoint distances.

The item in oru regarded as 'sure' is item 4 ("I fell and broke my leg"). The phrase "breaking one's leg" is taught during the 3rd year of junior high school and at the beginning of the 1st year of high school. Items 10 & 4 are chosen by the most subjects and consistently by the same subject. In terms of the sum of consistent and variable responses (criteria 2) and of consistent responses (criteria 3), items 10 and 4 are comparable. But according to the small scale investigation (appendix 4), the phrase "breaking one's leg" appears 6 times, but the combination of "break" with "the stem (stick) of an umbrella" does not appear in the textbooks. Thus, the 'sure' item is tentatively regarded as item 4.

The result concerning oru is summarized in Table 8.

oru ('sure' item: 4)

<table>
<thead>
<tr>
<th></th>
<th>consistent</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>p=0.3167 N.S.</td>
<td>p=0.0229 N.S.</td>
</tr>
<tr>
<td>DIM2</td>
<td>p=0.0833 N.S.</td>
<td>p=0.2167 N.S.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>consistent</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS-SIM</td>
<td>p=0.2333 N.S.</td>
<td>p=0.225 N.S.</td>
</tr>
</tbody>
</table>

(p stands for p.)

Table 8
No rank order correlations are significant, as in Analysis I. We have already seen that the items belonging to the conceptual category of 'folding' are not regarded as translatable. By comparing the judgement data with the MDS result, we can recognize that the two other categories are altogether rejected:

![Diagram]

The completely rejected clusters are circled.

Fig. 13

Removing the items which the subjects reject as a group, the rank order correlations are again computed:

oru (:'sure' item: 4)

<table>
<thead>
<tr>
<th>Asymmetric separable</th>
<th>consistent</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>DIM2</td>
<td>-0.3</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

| Integral MDS-SIM      | 0.6        | 0.55 N.S. |

Table 9

If the subjects categorize the items on the list and reject some categories as untranslatable, probably on the basis of semantic remoteness, the above data treatment can be justifiable. The statistical result reveals that as in kowasu and kudaku, DIM1 correlates positively with the judgement data and DIM2 correlates negatively. In this sense oru appears to support the joint process of assimilation and accommodation (see above). The statistically significant
correlation is reached under the assumption of integral processing. This may be an indication that the joint contribution of the two dimensions is reflected in the good correlation yielded by integral assumption.

As the MDS result shows graphically, there are 5 correlated pairs (1 & 6, 2 & 4, 3 & 5, 2 & 9 and 1 & 7) and 2 orthogonal pairs (((7 & 10) and (7 & 4)) and ((3 & 8) and (8 & 9))). However, only one correlated pair (3 & 5) shows the pattern predicted by separable processing. This result contrasts with the case of kudaku, where the two dimensions are separately attended to by the subjects and at the same time, the integral assumption yielded the high correlation with the judgement data. Whereas in the case of oru, there is not much indication of separable processing and the integral assumption yields a statistically significant but poorer correlation than kudaku. In this sense it appears that in relating separable and integral processing to the judgemental process, these two kinds of processing require careful examination. Although the usual definition of integral processing tends to exclude the possibility of separate attention to each dimension, the case of kudaku suggests that greater the indication of separable processing there is, the higher the correlation becomes under the integral assumption.

The item in yaburu regarded as 'sure' is item 8 ("Mr. Yamada broke his promise"). The phrase "breaking one's promise" is specially taught as an idiomatic phrase and is often chosen as a useful expression and an item for rote memory work. Item 10 ("I tore the umbrella") was accepted by the subjects as translatable using "break" as often as item 8. Also, in terms of the present consistency criterion, item 10 was identical to item 8. However, the NL sentence for item 10 clearly indicates a 'tear' concept rather than a 'break' concept in Japanese. The subject's acceptance is thus regarded as interlanguage speaker's message adjustment (a term from Váradi, Hatch, Tarone); message-wise a torn umbrella can be approximated to a broken one. The English word "tear"
is introduced in the 4th form of secondary school, since it is the policy of the Ministry of Education that irregular verbs (in terms of conjugation) should be taught from the 3rd form upward. The following are examples to show how an English 'tear' is introduced:

(1) In a few minutes, (a male wolf) had torn (a female wolf) to pieces.  
   (At the end of 4th form.)

(2) Suddenly, among the forest of hands that groped towards Richy, trying to tear off scraps of the clothing of their idol, ...  
   (At the end of 5th form.)

(3) She wore overalls and a torn skirt, and she was barefooted.  
   (At the end of 6th form.)

From (1) it is hard to associate it with the basic meaning of "tear". Recently the policy of the Ministry of Education has changed, but around the time this pilot study was run, it is fair to say that from the 4th form upward teaching centred on reading and writing; and reading materials were particularly likely to contain archaic or literary expressions which the students were encouraged to learn by heart for the purpose of translation and speedy reading. Consequently, as the present data show, "tearing an umbrella" and "tearing the skirt" in Japanese was considered by the subjects to be "breaking an umbrella" and "breaking the skirt" in English. From the viewpoint of interlanguage competence, item 10 (tearing the umbrella) is a good candidate for the 'sure' item. However, item 8 (breaking one's promise) is chosen equally often. On the ground that item 10 is a result of interlanguage message adjustment and communication strategy, which could be distinct from interlanguage competence, item 8 is regarded as 'sure'.
The following table summarizes the result concerning yaburu:

**Yaburu**

<table>
<thead>
<tr>
<th>Separable ('sure' item: 8)</th>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>p=0.75</td>
<td>p=-0.4736 N.S.</td>
</tr>
<tr>
<td>DIM2</td>
<td>p=0.4405 N.S.</td>
<td>p=0.6071</td>
</tr>
</tbody>
</table>

**Integral**

| MDS-SIM                   | p=0.8095   | p=-0.4286 N.S. |

(p stands for \( p \).)

Table 10

According to Table 10, DIM1 correlates positively with the consistent responses and negatively with the variable responses. DIM1 is dominant in terms of consistent responses, while in terms of variable responses DIM2 is dominant. In this sense the asymmetric separable assumption is supported by the data. However, as before, the integral assumption yields a high correlation. The MDS configuration shows that two pairs (8 & 1) and (3 & 5) are correlated. The pair (8 & 1) elicited the same number of consistent and variable responses respectively (see Table 2 yaburu in 1.1, Appendices). For this reason the prediction of separable processing in terms of the MDS configuration is born out by the judgement data with respect to this pair. But the pair (3 & 5) elicited, the different number of consistent responses (see Table 2 in 1.1, Appendices) and therefore in terms of consistent responses this pair does not prove the occurrence of separable processing. But in terms of variable responses, the pair (3 & 5) supports the separable assumption.

According to Table 10, the data as a whole supports the integral assumption, since the assumption yielded the better correlation with the judgement data than the separable assumption. In our discussion of oru data, we have parenthetically suggested that integral processing is compatible with separable processing, and that an equal amount of separate attention (therefore balanced attention) to each dimension can produce an integrated assessment. This speculation is not supported by the data, but as we have
seen above the data supports asymmetric separable processing. (This might suggest the possibility that integral processing includes asymmetric processing as its integral part. Chap. 9 shows that it is the case.)

In discussing kowasu and kudaku above, we have argued that the NL assimilation is reflected in positive correlations between the judgement data and the MDS dimensions derived from a set of NL items, and that the TL accommodation is reflected in negative correlations. This NL assimilation and TL accommodation is obtained in the correlations with DIM1 (see Table 10).

Some peculiar contrast may be noted between consistent and variable responses. This is distinct from the other verbs. In the other verbs, if one dimension shows a positive or negative correlation in terms of consistent responses, the variable responses follow the same pattern in the respective dimension. But in the case of yaburu the positive correlation of consistent responses is not replicated in the variable responses. That is, in the other three verbs, the NL assimilation (positive correlation) and TL accommodation (negative correlation) were observed in the dimensional relationships, i.e., one of the two dimensions takes either of the two processes of NL assimilation and TL accommodation. But as for yaburu the NL assimilation occurs in the consistent responses and the TL accommodation, in the variable responses.

Interlanguage (IL) study has regarded the systematic occurrence of any item as reflecting IL competence. Whereas, the variable occurrence has been regarded as relating to strategies of communication (e.g., "topic avoidance", "paraphrase", "appeals for the authority", "NL borrowing", etc.) and other factors of linguistic planning (e.g., a shift of topic and lexical search). Under communicative pressure the learner adopts various communication strategies to achieve his intended objective. It has been suggested that this pressure prevents the systematic occurrence of
IL competence and results in variable performance. Since this pilot study deals with experimentally elicited data, it is not strictly appropriate to relate the discussion to the distinction of IL competence and performance. But the present distinction of consistent vs variable responses may be compatible with the basic distinction between IL competence and performance (see §4.3.3). In view of this, we might be able to suggest that the NL assimilation associated with DIM1 reflects IL competence and that the TL accommodation associated with DIM1 occurs in IL performance.

By introducing this theoretical framework into the discussion, it is possible to give a plausible, although merely anecdotal explanation of the possible source of such IL expressions as "breaking the skirt" in the sense of "tearing the skirt" and "breaking an umbrella" meaning "tearing an umbrella". The discussion also provides us with a reason why item 10 ("breaking an umbrella") was another possible 'sure' item in a learner's IL.

I have indicated in this section that the confusion of "breaking" and "tearing" arises from the syllabus. The NL concept of yaburu (tearing) is first associated with the TL expression "breaking" (=yaburu) one's promise". This is introduced at the 4th (recently the 3rd) form of secondary school. Several months later, the TL word "tear" is introduced. But since the students are already 16/17 years old by then, the topics in textbooks are intended to match their intellectual level and curiosity, so that the basic use of "tearing" a piece of cloth or paper hardly occurs (see example sentences above). For example, in one high-school textbook "tear" first occurs as "tearing a flower" used in the sense of tearing the petals off. This example is translated into Japanese using chigiru (pluck off). Similarly, the previous example of "tearing a female wolf to pieces" is translated into "konagona-ni (to pieces) hikisaku (the composite verb of hiku (pull) and saku (split)". Neither is translated into Japanese with yaburu (tear). Although these three verbs can be brought together conceptually (see Table 1 appendix 3), their connection may
not be immediately recognizable to a learner, since each of the verbs has a distinct lexical field. If a student learns "tear" in association with plucking off petals or tearing flesh, it may be difficult for him to work out more basic uses: e.g., "tearing a skirt" (item 3) or "tearing off a page". On the other hand, "breaking one's promise" is translated into Japanese with "tear" (yaburu). It is possible for a learner to concretize the abstract use of 'break' and consequently to establish some conceptual equivalence between "break" in English and "tear" in Japanese at least in his IL competence. This is the result of confusing the abstract use of a TL expression with the concrete concept in NL. We will call this concretization as an abbreviation for the explanation below. This confusion is encouraged again by syllabus. The TL 'break' (kowasu) is first introduced with domestic utensils. This association of "break" (kowasu) with domestic items might lead to the association of another "break" (yaburu) with items belonging to the same category including skirts & umbrellas. If so, the syllabus encourages some concretization. This confusion is not peculiar to the present group of subjects. As we see in chap.10, the same concretization is observed in all groups. The subjects who took part in this pilot study are the most advanced of all groups participating and in this sense the phenomenon of concretization appears to be fossilized and to persist in IL competence among the Japanese learners of English dealt with here.

The above observation is relevant in that the data ought to be examined from the viewpoint of another sure 'item' (i.e., the fossilized item 5). This 'sure' item is clustered with item 3, according to the MDS result. In order to examine the possibility of an analogical process on the basis of 'sure' item 5, the distance measures along DIM1 can be used, since DIM1 represents a concrete/abstract dimension. In calculating a rank order correlation, the four items (5, 3, 6, 2) are considered: 6 is an obvious choice, since it is the 2nd closest to 5. 2 rather than 7 is chosen, since 2 is a metaphorical use derived from a body-part term.
Sure item (5)

<table>
<thead>
<tr>
<th></th>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>$p=-0.5$</td>
<td>$p=1$</td>
</tr>
<tr>
<td>DIM2</td>
<td>$p=-0.5$</td>
<td>$p=1$</td>
</tr>
</tbody>
</table>

(p stands for $\rho$.)

Thus, the analogical process based upon the result of concretization (item 5) clearly belongs to the domain of variable performance rather than IL competence, since the result yields a perfect correlation 1. However, there is an indication that IL competence relates to the 'sure' item 8 which was learned first. Because both DIM1 & DIM2 yield $\rho=-0.5$, it follows that in judging the translatability, the most dissimilar item from item 5 in terms of the two dimensions is intuited by the subject as the best candidate. The most distant item from item 5 is item 8 (breaking one's promise), according to the distance measures computed by MDS. This item 8 is the abstract 'sure' item and here yields negative correlation. Thus, we notice that the 'sure' item which was learned first functions as an accommodatory force in the analogical process based upon the derived 'concrete' 'sure' item.

The reverse situation should have occurred when item (8) was initially taught. The concrete item (5) is the remotest item from 8. Therefore when accommodation occurs, this concrete item 5 must have functioned as an accommodatory pull. This might offer a partial explanation of the source of concretization. It is worth noting that accommodatory forces prevent the continuation of the analogical process. In theory, without accommodatory forces and pulls the analogical process can recur continually. Furthermore, the two processes of assimilation and accommodation may contribute to regulating the judgemental process.

Then, the next question to be asked is how much effect the abstract 'sure' item has upon the other abstract items as the basis of the analogical process.
The question is relevant in that the 'sure' item 8 appears to play a major role in IL competence, as we have seen: (compare correlations in Table 10 with those in Table 11). According to the above analysis items 5, 3, 6 & 2 form a group. In order to see the effect of the originally taught abstract sure item 8, the best procedure is to calculate correlations by excluding the concrete items one by one so that the effect of concrete items can be gradually reduced. By doing so, the effect of the abstract 'sure' item will be revealed at each step of the calculation. Since items 5 & 3 do not differ in terms of their respective values along DIM1, it is decided to exclude them.

(1) Excluding item 3 & 5

<table>
<thead>
<tr>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>-0.0571</td>
</tr>
<tr>
<td>DIM2</td>
<td>0.7429</td>
</tr>
</tbody>
</table>

(2) Excluding items 3, 5 & 6

<table>
<thead>
<tr>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>0.05</td>
</tr>
<tr>
<td>DIM2</td>
<td>0.9625</td>
</tr>
</tbody>
</table>

(3) Excluding items 3, 5, 6 & 2

<table>
<thead>
<tr>
<th>Consistent</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM1</td>
<td>0.6</td>
</tr>
<tr>
<td>DIM2</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Table 12

Incidentally, it may be noted that the above 3-step procedure of gradually excluding concrete items (for the purpose of examining the effect of abstract 'sure' item 8) is justified
by the plausibility of asymmetric separable processing derived from DIM1. The results with DIM1 clearly indicate the increasing rate of abstract effect from procedure (1) where $\rho = -0.057$ to procedure (3) where $\rho = 0.6$. In procedure (3), all items belong to abstract or psychophysical items and therefore DIM1 practically amounts to the left half (i.e., abstract half) of the abstract/concrete dimension. The result obtained by procedure (3) is for those items belonging to the abstract part of dimension 1.

Since our current question is concerned with the effect of the abstract sure item, relevant results to be discussed are those concerning DIM2, DIM1 being more closely related to the sure item 8 (see the MDS configuration of yaburu). The three results in Table 12 indicate that the inclusion and exclusion of items 6 and 2 influence correlations. A turning point from the lower correlation to the higher correlation is determined by the presence or absence of item 2. Procedure (1) in which item 6 is included yields the lowest of all, whereas by excluding 6, item 2 becomes the remotest item from the 'sure' item 8 along DIM2. This procedure (2) yields the highest correlation of the three procedures. The further exclusion of item 2 in procedure (3) does not improve the correlation coefficient obtained by procedure (2). Thus, procedure (2) provides us with the best prediction of the observed judgemental data in terms of the analogical process (i.e., assimilation) based upon the abstract 'sure' item. In this context it will be recalled that the analogical process based upon the concretized 'sure' item (5) is best accounted for by the procedure presented in Table 11: $\rho = 1$.

Thus, by combining those procedures which yielded the best predictions, the following pattern emerges:

```
sure item (8)                      sure item (5)

DIM2                              DIM1
p=0.9625                           p=1

Fig.15

item 2: turning point
```
This also provides evidence to the effect that the two clusters consisting of abstract and concrete items are distinctively conceived by the subjects in their judgemental process.

Another interesting feature emerges from procedure (1) (see Table 11). A very small amount of accommodation is observed in DIM1, while DIM2 shows NL assimilation. This pattern coincided with the marked tendency observed among the other three verbs. This is the result which was not revealed when the cluster was not incorporated into the analysis (compare the results in Table 10 with those in Table 12.). This suggests again the importance of clustering in analyzing the judgemental data.

In terms of NL assimilation and TL accommodation, yaburu shows the most intricate relationship of all. Three relationships are recognized.

(1) According to the results (Table 12), DIM1 correlates positively with consistent responses and negatively with variable responses. As before, this feature was interpreted as reflecting NL assimilation and TL accommodation. Yaburu reflects the NL assimilation and TL accommodation in the consistent and variable responses respectively. Whereas in the other verbs, the NL assimilation occurs in one dimension and the TL accommodation, in the other dimension. This idiosyncrasy of yaburu was explained in relation to the distinction of IL competence and performance.

(2) The results in Table 11 provides the evidence that the analogical process involving the concrete 'sure' item 5 which was never taught as translatable at schools is related to the originally taught abstract 'sure' item 8 (see Table 11 and Fig. 14 in relation to the comments made.) This originally taught abstract 'sure' item 8 at first functions as an accommodatory force toward the TL. This TL accommodation was observed both with DIM1 and DIM2 with respect to the consistent responses of IL competence, establishing as a result the secondary concrete 'sure' item 5 in the mind of a learner. That is, although our first analysis (Table 10) did not show any trace of accommodation
in the IL competence, by categorizing the data in terms of the secondary 'sure' item 5, the result in Table 11 reveals evidence that the TL accommodation was in operation.

(3) Table 12 indicates that there was a small amount of accommodation in DIM1, but that DIM2 produced the NL assimilation (see procedure (1) in Table 12). This is the usual pattern found in the other three verbs, i.e., one of the two dimension is subject to the NL assimilation and the other dimension, to the TL accommodation. This finding suggests that the importance of categorization effects in analysing translatability judgement data.

It may be noted additionally that Piaget's notion of assimilation and accommodation has been regarded as maintaining "cognitive equilibrium". In the present analysis we have seen that the NL assimilation revealed as the positive correlations with one of the NL dimensions is counter-balanced by the opposite force of TL accommodation which is revealed by the negative correlations with the other dimension. When we think of these opposite forces in the two-dimensional NL spaces, we can see that these counter-active forces contributes to the maintainance of equilibrium in the NL spaces.

§ 4.4.3 Analysis III

This analysis adopts Torgerson's measurement of categorical judgement which closely follows Thurstone's method of paired comparison. The purpose is to compare the interpoint distances computed by MDS (MDS-SIM) with Torgerson's (SIM) and to decide which provides a better analytical tool for our present purpose of analyzing the judgement data. As indicated previously, it is always possible for the subject to adopt a primitive form of similarity judgement, i.e., paired comparison. As analyses I & II clearly indicate, MDS provides us with a very useful way of looking at the judgemental process. It is largely because the group of subjects are advanced learners (15-23 years' exposure to English) and are expected to possess well-structured NL interlexical knowledge. However, 5 groups of subjects in the main study are less
advanced (2-10 years' exposure to English) and their NL interlexical knowledge might not be as well-structured as those subjects who provided the input data for MDS (late-twenties to well over 50 years old). While MDS requires inherently well-structured input, SIM is more versatile in this respect. Thus, this analysis has a practical purpose: i.e., if SIM is comparable to MDS-SIM and the younger subjects adopt a paired comparison, then Torgerson's method is still available to us. Besides, SIM does not make any specific assumption concerning a type of processing the subjects adopt. In other words, whether integral or separable SIM can be used. For this pilot study the simplest condition (IV) is used. Because of this, we can only compare MDS-SIM with SIM. Since the number of subjects available for the translatability test was very limited (15 Japanese at Edinburgh), the merit of Torgerson's equation was not made full use of here, but this method provides a useful way of predicting the "inductive (analogical) limit" (see chap. 7 particularly §7.4 & chap. 10).

Results
The results of SIMs calculated on the basis of Torgerson's method are given in 1.2, Appendices. The assessment of 'sure' items for each verb follows Analysis II. The following table 13 lists rank order correlations of the MDS-SIMs with the judgement data and of the SIMs with the judgement data. In oru and yaburu, correlations are also calculated separately in view of the categorization effects.

<table>
<thead>
<tr>
<th>'sure' item</th>
<th>method</th>
<th>consistent</th>
<th>variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kowasu</td>
<td>4/8</td>
<td>MDS-SIM</td>
<td>0.3393</td>
</tr>
<tr>
<td></td>
<td>4/8</td>
<td>SIM</td>
<td>0.9107</td>
</tr>
<tr>
<td>Kudaku</td>
<td>4</td>
<td>MDS-SIM</td>
<td>0.8333</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>SIM</td>
<td>0.8691</td>
</tr>
<tr>
<td>Oru a</td>
<td>4</td>
<td>MDS-SIM</td>
<td>0.2333</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>SIM</td>
<td>0.4667</td>
</tr>
<tr>
<td>Oru b + categorization (see Analysis II)</td>
<td>4</td>
<td>MDS-SIM</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>SIM</td>
<td>1.0</td>
</tr>
</tbody>
</table>

149
| 'sure' item method consistent variable |
|---------------|--------|----------------|
| Yaburu_a      | 8      | MDS-SIM 0.8095 -0.4286 |
|               | 8      | SIM 0.5595 -0.3571     |
| Yaburu_b (excluding items 3 & 5: see procedure (1) in Table 12) | 8 | MDS-SIM 0.2571 0.2286 |
|               | 8      | SIM 0.3143 0.2571     |
| Yaburu_c (excluding items 3,5 & 6: see procedure (2) in Table 12) | 8 | MDS-SIM 0.7 0.75 |
|               | 8      | SIM 0.6 0.7           |
| Yaburu_d (concrete group): see Table 11. | 5 | MDS-SIM -0.5 1.0 |
|               | 5      | SIM -1.0 0.5          |
|               |        | Table 13              |

The above result on the whole suggests that, apart from kowasu, yaburu_a and yaburu_d, MDS-SIM and SIM are comparable irrespective of data-treatments (± categorization) and the kinds of verbs. This implies that if the NL interlexical relationship is not well-structured in future data, the 'feel' of similarity can be measured by SIM.

Since there is another opportunity to describe SIM in chap. 5 & 7, only a few characteristics are mentioned here. The way the present study makes use of SIM assumes a bivariate normal distribution:

![Graph](image_url)

Each normal distribution represents one pair and the red
line, a similarity continuum. The central tendency (i.e., mean) of each normal distribution on the similarity continuum is the value representing the 'feel' of similarity between the pair. Although each word has various semantic features and internal relationships, the method assumes that only two variables (x, y) are in operation. It does not make any specific assumption regarding what these x and y stand for. X and y may be the result of global assessment made by the subject on the basis of various semantic features, or it could be the result of one dominant semantic feature operating in the process of paired comparison. However, the method is specific about one assumption: that is, there is at least one commonality represented by f (x, y) which is general enough to cover all items in the list. In this sense, it is a useful method for a set of polysemous words. Unless there is some conceptual similarity among the items, the application of this method is not tenable. The results of SIM produce a linear scale of similarity continuum along which each pair is arranged in the order of similarity. Due to the specific assumption of one commonality, the method corresponds to integral processing. However, the source of this commonality could be the result of separable processing or asymmetric separable processing on the part of a learner. We are not certain about this. We are only sure of the commonality and the presence of the strategy of paired comparison on the part of the subject's judgemental process. Since paired comparison is a basic unit of analogical process, it would be useful to quantify the primitive mode of analogical process in dealing with less advanced learners below.

Whereas MDS is a multi-dimensional scale, it presupposes the presence of well-defined structure on the part of the subjects. It also presupposes that well-differentiated relationships have continuous features (as opposed to discrete semantic features) and that these continuous features (which are assumed as functioning as criteria in making similarity judgements) are converted into dimensions (see chap. 5 & 9). In this respect, MDS is useful on the condition that the subject
can be assessed as possessing a sense of well-defined structure for a set of items. MDS is efficacious when the subjects have linguistic sensitivities and are capable of reflecting these sensitivities in their numerical responses of rating.

According to Analysis II, apparent experimental confirmation of asymmetrical processing at the superficial level (except for vaburu) is deceptive in that the subjects were in fact separately attending to the two dimensions, adopting an analogical (similarity-based) strategy with respect to one dimension and a dissimilarity-based strategy with respect to the other dimension. The former is a manifestation of NL assimilation and the latter, TL accommodation. These similarity-based or dissimilarity-based strategies are reflected in positive or negative correlations respectively. There is a degree of dominance, one dimension dominating over the other, but it is significant to bear in mind that the dominant dimension cancels out in terms of statistics the cognitive presence of the other dimension. In this sense, Kellerman's core hypothesis may be a simplistic account of the judgemental process.

At this point, it may be recalled that the integral assumption often yields correlation coefficients which are comparable to or only slightly poorer than those with the dominant dimension. This observation leads us to postulate two possibilities in interpreting the correlations with the MDS-SIMs.

The high correlation under the integral assumption can be mean either or both of the following:

1. It reflects the over-all assessment of the relevant dimensions. The separate and balanced attention to these dimensions can be assumed, involving some analysis on the part of the subjects. The MDS results must show the presence of well-defined interlexical structure on the part of the subjects.

2. Due to the nature of the task, the impressionistic
global assessment is sufficient. The undifferentiated intuitive grasp is called for.

The second type of integral processing resembles SIM. This could account for why apart from kowasu, yaburu_a and yaburu_d (Table 13) MDS-SIM and SIM yield comparable results up to a point. Yaburu_a does not appear to involve much of paired comparison, since yaburu_b, yaburu_c and yaburu_d clearly indicate that the subjects are categorizing the list of items. It is not paired comparison but categorization that is a major factor here. This reflects the poorer correlation with SIM (p=0.5595) but the higher correlation with MDS-SIM (p=0.895). On the other hand, in kowasu SIM yields the highest correlation of all (see Analysis I Table 2 and analysis II Table 4). This can be taken to indicate that paired comparison is the strongest strategy used by the subjects.

§4.5 Summary

In this pilot study the three kinds of methods are used to examine Kellerman’s core hypothesis. In terms of the major types of processing, Kellerman’s hypothesis assumes asymmetric separable processing. Analysis I assumes asymmetric separable processing, as in Kellerman. Two verbs out of the 4 support Kellerman’s position. Since this result could be due to the different nature of verbs, the verbs are analyzed in terms of co-occurrences of category types. There are not immediately apparent differences in this respect. The discussion indicates the importance of the clusters produced by MDS, particularly when we assume that the MDS result is an adequate representation of NL interlexical structure. It is argued that in the two verbs which yielded poor correlations the effect of categorization is distinctly observed. In this sense, Kellerman’s core hypothesis (which heavily relies on one specific dimension and assumes a simple correspondence between this dimension and the translatability judgement) needs to be modified.
Analysis II incorporates the idea of 'sure' item, along with the two types of processing: asymmetric separable and integral. It is postulated that there is a 'sure' item which the subject is confident of (see the operational definition of 'sure' item (see §4.2)) and that the subject intuits a 'feel' of similarity between the 'sure' item and the rest of the list. This provides the basis for analogical process. The results give us a more realistic understanding of the judgemental process. It is shown that the judgement of translatability is derived from the joint operation of NL assimilation and TL accommodation. This pattern emerges without considering the effect of categorizations in kowasu and kudaku; whereas oru and yaburu call for the effects of categorization, where the same interactive operation of NL assimilation and TL accommodation was clearly observed. The discussion leads to the observation that the apparent superiority of one dimension over the other is deceptive, since the close examination of the judgement data correlated with the MDS results reveals that both dimensions are separately well attended to by the subjects; one dimension provides them with a similarity based analogical strategy (NL assimilation) and the other counteracting dimension provides the basis for a dissimilarity-based strategy (TL accommodation). It is argued that this mechanism might account for regulated behaviours in NL transfers which are not particularly obvious at first glance.

The assumption of integral processing which yields overall assessment of similarity produces statistically significant correlations with the judgement data. The correlations are often as good as those under the assumption of asymmetric separable processing (i.e., the correlation with the superior dimension). This is interpreted as supporting the joint operation of two dimensions in the judgemental process. Thus, there are two ways in which integral processing can manifest itself: (1) overall assessment involving analytical understanding of relevant features; (2) unanalytical impressionistic 'feel' or global grasp (see §4.4.2 and §4.4.3). As for (1), MDS is useful and for (2) SIM is useful.
Analysis III examines whether the subjects adopt in a judgmental process a strategy of paired comparison (SIM) between the 'sure' item and the rest of the list. The subjects who took part in this pilot experiment were advanced learners with developed NL knowledge in which an NL interlexical structure can be assumed to be well-structured. Thus, although kowasu offers evidence that paired comparison was adopted by the subjects, the other three verbs show that MDS-SIMs are slightly superior to SIMs, reflecting that MDS is suitable for the representation of well-defined NL structure.

§4.6 Plan of research

Before one can cope with developmental data, the following preliminary investigations are regarded as necessary.

(1) In the above, the three types of processing were considered, but they are put forward as assumptions. The judgement data were analysed according to the specific assumptions of processing. This is not satisfactory. When we can determine what form of processing is characteristically adopted by the subjects in the process of making similarity judgements prior to the analysis of the judgement data, this is more satisfactory. MDS is again useful in specifying a form of processing the subject adopts, according to several psychologists (Arnold, 1971, Shepard, 1974 and Garner, 1976). This is discussed in chapter 9.

(2) Analysis I shows the importance of categories, i.e., which notional category a given item belongs to determines clusters in the MDS results. Thus, it would be more appropriate to use learner-generated items rather than items chosen arbitrarily by the experimenter. A small scale experiment is run to collect learner-generated instances belonging to some notional categories (chapter 5).

(3) The 5-point scale for rating in this pilot study follows Kellerman (1977). The small experiment is run to determine the satisfactory number of steps. This experiment
is motivated by another reason. In applying Torgerson's method, the number of steps becomes more important. As Tables 1 - 4, 1.1, Appendices show, there are a few empty cells in which only frequency is written. This is firstly because the 5th step of the present rating scale is not enough for the subjects to make a finer differentiation. Secondly, because the special care need to be taken to collect all possible paired ratings. The point is discussed in chap.5.

(4) Since this pilot study shows the importance of subjective organization, a free recall experiment is run to test this point. This experiment is to see whether or not two-item groupings are observed so that they can provide some plausibility in using a method of measurement based upon paired comparisons (see chap.6).

(5) Furthermore, in using SIM the method needs to be attested by separate experiments. Unless we can experimentally demonstrate that numerical representations have some psychological plausibility, the analysis of data based upon the derived measures might sound less convincing (chapter 7). In addition, it seems that SIM can compute the judgemental limit. Chap.7 aims to examine whether the derived judgemental limits are accurate enough.

(6) This pilot study shows that NL assimilation and TL accommodation relate to the judgement of translatability. The NL interlexical structure might go through some change in the course of learning English if the amount of TL accommodation is considerable. The issue relates to the classical distinction between coordinate and compound bilinguals. Although NL interlexical relationships are examined by MDS and SIM separately according to each of the 5 groups, the issue seems to be still relevant in that, if separate meaning systems for TL and NL (coordinate bilinguals) are developed, the question remains over what it means for us to regard some thing as translatable. Implicit in the notion of NL transfer is that concepts remain the same and that a learner
relexicalizes a concept in using TL or NL. For this reason, it is important for us to see whether Japanese learners of English are compound or coordinate bilinguals (chapter 8).

(7) This pilot study has dealt with 4 verbs: kowasu, kudaku, oru and yaburu. In relating NL to the judgement of translatability the former two behaved rather similarly to each other and the latter two, to each other. kowasu and yaburu were chosen belows. There is a pedagogical reason for this choice. According to standard textbooks in Japan, English "break" with respect to kowasu is initially taught with concrete nouns as subjects or objects, while yaburu (tear) with respect to "break" is taught with abstract nouns (e.g. promise) before they learn English "tear". As we have seen, this syllabus might be the source of such common errors as "breaking paper" among Japanese.

(8) Because of free recall experiments (chapter 4) and two experiments involving reaction times (chapter 7 & 8), the sentences to be presented below are shorter than those used in this pilot study and the simple combination of verbs and object nouns is used.
Chapter 5  Pilot Study (2): materials, rating & MDS elicitation

§5.0 Purpose

This pilot study (2) proposes to determine (1) object nouns which serve as experimental items, (2) the number of steps in the rating scale and (3) the method of eliciting input data for MDS. These preliminary investigations are necessary, since the IL study aims to research learner language and the more learner-centred the approach is, the less the risk of arbitrary decision-making on the part of the experimenter. Experiments concerning the above three objectives are reported below.

§5.1 The method of choosing object nouns

The method of data collection to determine object nouns follows closely what psychologists have called "the production method" for collecting category norms (see §5.1.1.). This method determines representative instances, based upon learner-produced data. The method accords with the IL aim of a learner-centred approach. In the present situation, we present learners with verbs rather than category names, in order to see what nouns are most often produced by learners and which nouns collocate with the two verbs kowasu and yaburu in their production data. The procedure for this production method is simply to present a questionnaire of the form: _____ wo kowasu; _____ wo yaburu. The subjects are asked to fill in any object noun which occurs to them.

In collecting collocational data, I have an additional purpose in mind. In order to run a well-controlled experiment for free recall (chap. 6) and to record reaction times (chap. 7), both "objective and subjective familiarity" (see below) must be assessed. Psychologists have shown that among other things the following 5 factors bias the rate of learning, reaction times, and the amount of both recall and recognition: (1) association values, (2) imagery values, (3) the degree of meaningfulness, (4) frequency
count and (5) "conjoint frequency" (references in chap. 6).

As Umemoto makes clear (1969a), these biasing factors can be classified into the two major categories: subjective and objective familiarity. The number of associations and the amount of imagery each word arouses relates to the degree of subjective familiarity. The frequency count is normally obtained by counting the number of occurrences of words in journals and books, and the conjoint frequency, by a frequency count of instances relative to a category norm. These reflect objectively assessed familiarity. The degree of meaningfulness may be regarded as somewhere between objective and subjective familiarity. The idea of meaningfulness usually applies to artificially coined "nonsense" words and relates to the ease of pronunciation and conceptualization. This pilot study is not concerned with the meaningfulness value, since the materials are from a meaningful natural language.

In the above, the frequency count is viewed as reflecting the degree of familiarity. That is, if a given item occurs frequently in daily use, the subject is considered to be familiar with the item. Although the frequency count provides an objective assessment of familiarity, it is nonetheless an indirect measure of an individual's familiarity. In comparison with this indirect measure, psychologists' methods (of counting or the use of a rating scale) involving associations and imagery are directly concerned with a part of what is internalized in cognition. The notion of subjective familiarity (if it includes what is internalized as the knowledge of language) might suggest something more than associations and high/low imagery. The psychological notion of subjective familiarity appears to be relevant to the IL study in view of the distinction between passive and active knowledge of NL or TL.

The subjective familiarity can be seen as relating to the active part of linguistic knowledge. How far an individual feels familiar with the item (i.e. 'sure' item) may be
the outcome of (1) good learning (as opposed to poor) and (2) the frequent use of this specific knowledge. These factors together suggest that highly familiar and 'sure' TL items are those which are internalized as a part of the IL in a highly available form. It would appear that unless the knowledge is readily available rather than merely passively accessible, it is not a part of active knowledge. The frequent use of this available knowledge may mean that it has been activated sufficiently, and that the degree of familiarity (i.e. sureness) is consequently strengthened. Once the knowledge is thus consolidated, it spreads a network of connections and associations among various knowledge sources. In this sense, the evaluation of associations and imagery is an effective means of assessing the development of not only subjective familiarity but also the associative network of vocabulary.

Furthermore, once the "production" method is used, as in previous psychological experiments, it usefully distinguishes what is actively available (active knowledge) from what is merely accessible (passive knowledge)—see §9.1. In this sense we can regard the "associations" produced by a number of subjects as evidence for the extent of availability and of a well-developed associative network. If the highly developed associative network is available to the subject, it will be easy for him to produce many associations in response to the presentation of a word.

In the production method, the experimenter normally ranks words produced by the subjects. This data-treatment is relevant to the present IL study. When words are ranked in the order of their increasing number of production, we may regard ranking as reflecting the line between highly active knowledge and merely passive knowledge. For this reason this psychological technique is adopted here.

One drawback of the production method for the present purpose is that the method does not account for why some words excite more associations than others or why some words

160
are remembered more easily than others. The words with high imagery and high association values are known to be learned more readily (see chap. 6), probably because imagery and association help to consolidate a learner's memory. It is also well known that the more often a learner is exposed to the same word, the more likely he is to remember the word (and hence such frequently exposed words tend to be produced more often by the subject if the experimenter adopts the production method). In this sense the objective frequency count appears to reflect an individual's familiarity with each word. Even though an item occurs frequently in daily life, it does not necessarily mean that the item is familiar to a specific person. But experimental evidence demonstrates that there is a significant effect of frequency counts in memory (Brown, 1976). In this sense, the degree of objective familiarity interacts intimately with the learner's subjective familiarity. Good interaction may lead to good learning.

Thus, it is appropriate to take into account two measures of objective and subjective familiarity: i.e. frequency counts (in this chapter) and association values (presented in chap. 6). In order to complement these two, one intermediary measure is considered here. It is basically a straightforward familiarity rating, but this rating is done not by the subject who produces a given object noun but by the other subject. The method is a sort of double-check. As mentioned above, the main objective of this section is to collect object nouns which co-occur with the two verbs, and to select some according to other external restrictions to meet the needs of some experimental tasks (chap. 6 & 7) so that biasing factors are kept constant in the future experiments.

§5.1.1 experiment

Three groups took part in the experiment: (1) 116 2nd year university students; (2) 122 2nd year high school students; (3) 80 3rd year junior high school students. They were presented with a questionnaire in the following form:
wo kowasu; wo yaburu. They were instructed to fill in the blanks. A 7-point scale for familiarity indices was presented next to each incomplete sentence. They were asked to write down as many nouns as possible, spending 30 seconds per verb: the idea of allowing 30 seconds per verb comes from Batlig & Montague (1969). Then, each subject exchanged the sheet with the subject sitting next to him and rated sentences in the sheet exchanged, using the 7-point scale from (1) very familiar to (7) very rare. Half of the subjects dealt with kowasu first followed by yaburu and vice versa.

§5.1.2 result

Table 1 shows the total number of nouns produced in the three age-groups. In kowasu, there is a significant difference among them ($\chi^2=3.32 \text{ N.S. (0.2}<P<0.1)$). This difference between the two verbs may be related to the fact that kowasu is the more general verb which can take as an object a variety of nouns, while the lexical field of yaburu is by comparison more limited. Thus, yaburu is less likely to reflect the richer NL vocabulary among the oldest group than kowasu does.

<table>
<thead>
<tr>
<th></th>
<th>Junior High</th>
<th>High School</th>
<th>University</th>
</tr>
</thead>
<tbody>
<tr>
<td>kowasu</td>
<td>97</td>
<td>122</td>
<td>186</td>
</tr>
<tr>
<td>yaburu</td>
<td>87</td>
<td>120</td>
<td>111</td>
</tr>
</tbody>
</table>

For the purpose of the present study, desirable experimental materials need to satisfy the following objectives:

(1) The items ought to range from literal to highly metaphorical uses of the two verbs. The translatability test requires the same items among the three age groups.

(2) The more typical the experimental items are,
the more general results will be in the main experiments.

(3) The free recall experiment below (chap. 6) requires materials whose frequency counts are constant for all the items on the list.

(4) Likewise, the free recall experiment requires materials whose subjective familiarity ratings are on average comparable to each other.

(5) Even for the main experiments if both the frequency counts and the subjective familiarity ratings are kept constant across the three age groups, the experiments can be said to be controlled. Another desirable consequence is that the NL lists will not contain additional variables as frequency effects and the effects of subjective familiarity. When the judgemental process is examined without involving these additional factors, the experiment can claim that it is largely about the judgemental process. If the NL lists contain additional factors, one can never be sure whether the experimental results are caused by these extra means factors or not.

(6) Experiment using a tachistoscope (chap. 7) requires some precision (0.4 second). Even a slight biasing factor can invite undesirable effects on reaction times. For this reason, well-known psychological facts concerning objective and subjective familiarity must be kept constant across the items on the list.

The above 6 objectives may be reduced to three operational conditions. They are listed in order of importance.

(A) Varied-list condition:
The items on the list must be varied from literal to metaphorical uses, since we are concerned with the polysemous nature of the verbs.

(B) Typicality condition:
Typical items are more desirable, and the frequency counts across both the items on the list and the three groups must be kept constant.
 Subjective familiarity condition:

When this index is kept constant either across the items on the list or across the three groups, this condition is satisfied. But if this is not feasible, we will run another experiment using another standard method to attest this point. Since the present experiment is based upon the production method and is basically concerned with counting the items produced by the subjects, the frequency counts have priority over the subjective familiarity rating.

Regarding the above condition (B), the frequency count can serve as an index of how typical an item is, since the more subjects who produce the same item, the more typical it is. Thus, the items produced by the subjects are ranked in the order of frequency from the most to the least frequent item (see Appendices).

As the following table shows, the rank order of typicality varies across the three age groups. The variation is more marked in kowasu than in vaburu. This indicates again that kowasu is a more general verb and that the developmental growth of NL vocabulary appears to be revealed in the varied responses over the three age groups. Thus, the typicality order in kowasu is liable to alter in the course of developing vocabulary.

(Tables 2 & 3 (see Appendices) list the ranks of high frequency words.)

The rank order correlations are not calculated, since some items are not produced across the three groups. The marked variation is however obvious in kowasu: e.g. "bridge" is ranked as 7.5 in the junior high school group, 17.5 in the high school and 67 in the university group. In vaburu the rankings are less varied: the most varied one in Table 3 is "record": i.e., 10, 18.5 and 11.5 in the respective groups.
As the above tables illustrate, among high frequency words there are not so many abstract nouns: in kowasu only one and in yaburu four abstract items. But these words are not equivalent in terms of frequency counts (see Tables 4-7, Appendices). Besides, although one body part noun "membrane" appeared in the two groups, it is not observed in the other group. Thus, it is clear that, in order to meet both the varied-list condition and the typicality condition mentioned above, one cannot make use of high-frequency words as experimental items, 2-frequency words and one-frequency words involve more items which meet the varied-list condition. This suggests that the operational condition C ought to be abandoned. The choice of experimental items from the low frequency nouns means that any generalizable observation concerning subjective familiarity ratings can not be made. This is entirely unexpected, since it was carelessly assumed that any native speaker irrespective of age would produce many typical nouns, being presented with such familiar verbs as kowasu and yaburu.

Tables 4 & 5 (in Appendices) report raw frequency scores and percentages for the nouns which are produced by more than 7.4% of each sampled population. For the less frequent words, the tables list the number of nouns which are tabulated according to the frequency counts: e.g., there are 44 1-frequency words and 23 2-frequency words. The tables illustrate that there are more words available for us to choose from in a decreasing order of frequency counts. We have seen in the above that the experimental items are to be chosen from the low-frequency words, due to the first two conditions A & B (see above). Although there are a greater number of words in the class of 1-frequency than in that of 2-frequency, the latter is chosen. It is mainly because 1-frequency items may be highly idiosyncratic or even deviant.

Tables 8 & 9 (in Appendices) list 2-frequency items according to the three age groups. The 11 words which are observed in all three groups are listed toward the top.
These 11 words are selected as experimental items, since they occur equally in all three groups and they belong to the same class in terms of the frequency count. Moreover, they form a mixed list. Thus, the 11 words satisfy the operational conditions A & B°.

The 11 words were written by the subjects in Chinese characters (not in Japanese letters). Chinese characters are more convenient (1) for disambiguation and (2) for a tachistoscope experiment (most items can be written in 2 Chinese letters, while in the Japanese letters the number of letters varies according to the number of syllables contained in a word.) Several school dictionaries were consulted to ensure that they were obligatory ideographs (characters) to be taught by the end of compulsory junior high school education. Although they are all obligatory characters, I was advised to remove one item "fun'iki" (atmosphere) by teachers at Shonan Middle School to which one experimental group belonged, because the subjects had not been taught one of the Chinese characters in "fun'iki" at the time of the experiment. Thus, the items on the yaburu list were adjusted to 10 sentences for this group.

§5.2 the choice of steps in the rating scale

We have seen in chap. 4 that Torgerson's method in conjunction with Thurstone's is useful (even in the simplest condition IV) in accounting for the analogical process of integral processing. We anticipate that at least the youngest group (3rd year junior high school students) will adopt this integral processing, jointly with the paired comparison as the simplest form of analogical process. If this is the case, Torgerson's method will be most useful. In order to make effective use of Torgerson's method, the number of steps on a similarity rating scale need to be properly determined. This is because the method is not merely exploratory, but can also be used to predict the analogical process on the assumption of paired comparison (see chap. 10). Also, the method is geared to estimate the magnitude between the steps, making the number of steps
crucial.

The Thurstonian method requires one extra step in calculating the magnitude between the consecutive steps. For instance, the following scale with 3 steps yields 2 category boundaries \( T_{g1} \) and \( T_{g2} \):

\[
\begin{array}{c}
\text{same} \quad \text{similar} \quad \text{different} \\
\ldots \quad \ldots \\
T_{g1} \quad T_{g2}
\end{array}
\]

Category boundary \( T_{g1} \) is the point at which category 'same' and category 'similar' meet each other. Category boundary \( T_{g2} \) is the point at which category 'similar' and category 'different' converge. Since the Thurstonian scale assumes a continuum, category 'same' continues infinitely \((-\infty)\) and the only point at which this category ends is specified as \( T_{g1} \). Similarly, category 'different' continues infinitely \((+\infty)\), but the point at which this category starts is specified as \( T_{g2} \). In this way, the only one magnitude between \( T_{g1} \) and \( T_{g2} \) is properly estimated. The more magnitude estimations we obtain, the more useful the outcome is for our purpose. The following experiments are to discover the appropriate number of category boundaries \( (T_g) \). The experimental items used for the paired comparison task are those obtained in the experiment above (see 5.1.2).

We use adjectives 'identical' and 'very different' to indicate the end points of the scale. Since we need one extra step, the minimal number of steps is 3. But this scale is too coarse, since it yields only one properly estimated magnitude. The customary (non-Thurstonian type) scales have 5 or 7 steps (e.g., Osgood's semantic differential (1975) uses 7 steps and Kellerman's study uses 5 steps). This study is based on the Thurstonian method. The category boundary \( (T_{g1}) \) in the Thurstonian method corresponds to the steps in the non-Thurstonian method. Thus, the scale of 6 steps (which yields 5 \( T_g \)s and 4 proper estimations of magnitude) may be adequate.
to begin with. The finer the scale is, the more information we will obtain. If we could make comparisons among the scales of $5T_g$, $7T_g$ and $9T_g$, it would be ideal. One can easily see that a scale of $11T_g$ would be implausible, since it is useless to have a scale finer than the subjects' discriminative ability and ratings obtained from overstrained subjects are less reliable. Because of the time factor (10 minutes are allocated), it was not possible to make comparisons among the three scales of $5$, $7$ & $9T_g$s; due to the educational principles of the institution, 10 minutes is considered the maximum for any non-educational use. In this pilot study, the scale of $5T_g$ is compared with the scale of $9T_g$ (which yields 8 estimates of magnitude). Under the present circumstances, the comparison between $5T_g$ and $9T_g$ may be more revealing than the comparison between $5T_g$ and $7T_g$, since the scale of $9T_g$ is twice as fine as that of $5T_g$.

An operational definition of a reliable and sensitive scale is as follows:

1. Inter-rater correlation (group consistency): i.e., whether or not many subjects show an agreement along one of the two scales on either of the two occasions. (test and retest).

2. Within-rater correlation (personal consistency): i.e., whether or not the same scale elicits more consistent answers from each subject on these two occasions.

3. Sensitivity: i.e., whether or not the scale is more effective in revealing significant differences among the paired sentences.

Since our informal definition of "reliable scale" may be rephrased in terms of the correlation between two equivalent sets of measures, we could either make use of two equivalent tests or repeat the same test after a certain period of time. The latter test-and-re-test format is chosen here, since it is difficult to compose truly "equivalent" tests for the present type of study. Although we know the frequency count of kowasu, other biasing factors
are not known to us (and we investigate only one more biasing factor, that of association values, in Chapter 6). An investigator may feel that 'breaking the heart' and 'breaking the spirit' are very similar, but there is no convincing ground for saying that these will be 'processed' as equivalent by his subjects. Thus, if the "equivalent" test items are not actually equivalent, they may cause some variation in the subject's ability to rate. The safest procedure seems to be a test and re-test format.

In this test and re-test situation, the possibility of remembering the previous test on the re-test is a major disadvantage. It is considered that this possibility is likely to be in proportion to (1) the time interval between the two tests and (2) the memorability related to the nature of the test items. As for (2), we have seen in 5.1 that subjective and objective familiarity are the major determinants of "learning" (i.e., "episodic" memory according to Tulving's well-known distinction between episodic and generative types of memory. That is, memory of details of a specific event such as the items in a test). As we see below (chapter 6), the association values concerning the experimental list are fairly homogeneous\(^5\). We have also seen in §5.1.2 that the experimental items are all 2-frequency items. As far as these two measures of familiarity are concerned, factor (2) mentioned above is fairly well controlled in this experiment.

Factor (1) is considered here in reference to Osgood et. al (1957). Using a test and re-test method for their semantic differential, they indicate that a one-day interval shows a minor change in their subjects' ratings. It suggests a possibility that the subjects might have remembered their own ratings. However, their study shows a sharp forgetting curve from a one-week interval to a three-week interval (1957:134). Thus, the interval of three weeks should be sufficient to allow for the greatest amount of forgetting on the part of the subjects.

Thus, the first interval between test and re-test was deter-
mined as three weeks. The time allocated to administer the experiment was only 10 minutes. There was only one class of students available for the experiment. This meant that the tests had to be repeated again (see Fig. 2). In this case, it is likely that the students would consolidate their "episodic" memory. Thus, the intervals are made longer, in order to prevent the consolidation of memory. Since 36 subjects were available, they were divided into two groups A & B. Group A started with a scale of $5_T$ and Group B, with $9_T$.

What one does not wish to happen is that the subjects recall their ratings in relation to the paired sentences. This is highly unlikely. They may recognize the paired sentences, but recalling the rating with respect to a specific pair would be difficult when their memory is partially faded. (It has been recognized that a slight memory is sufficient for recognition, but recall requires a more stable memory: see Kintsch, 1970b.)

The list of kowasu consists of 11 sentences, as decided in §5.1.2. Ideally all items on the list ought to be used for this experiment. A paired comparison consisting of 11 sentences means that the subjects compare 55 pairs. Since roughly 10 minutes was available, it was not possible to use all 55 paired sentences. Within this brief period of time, the subjects have to complete the following task. First, in order to understand the instructions in relation to the text of the test, they read the instructions and the text three times. Before they rated, they read the instructions once more. They may not follow these instructions completely, but this gives one a rough idea of how long the task takes, including the time spent on judging.
During the experiment, the subjects read silently. Inner speech is reported to cover 3 to 6 items per second: Norman (1969: 100). Since inner speech relates to rehearsal and chunking of incoming information, the unit of inner speech cannot be readily equated with a syllable. But, a syllable is sometimes regarded as a unit (ibid.). Thus, the number of syllables are counted under the following five conditions:

<table>
<thead>
<tr>
<th>No. of syllables</th>
<th>6 pairs</th>
<th>7 pairs</th>
<th>8 pairs</th>
<th>9 pairs</th>
<th>10 pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of scanning</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>1302</td>
<td>1518</td>
<td>1760</td>
<td>2028</td>
<td>2322</td>
</tr>
<tr>
<td>Instruction:</td>
<td>132 syllables per scan. 4scans=528syllables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range (fast performer — slow performer of inner speech)</td>
<td>217-434</td>
<td>253-506</td>
<td>293-587</td>
<td>338-676</td>
<td>387-774</td>
</tr>
</tbody>
</table>

Since the total time required should be less than 600 seconds, 7 pairs were chosen, which gives us roughly 4 minutes to deliver the experimental sheets. The first item (scale model) in Table 8, Appendices was used as a standard sentence with which the other 7 sentences were paired. That is, the 8 sentences from the top of the Table were used for this experiment.

Since categorial sorting (§5.3) is too elaborate to be used in a 10-minute test and re-test format, the method of rating pre-sorted sentences was used. In order to incorporate an aspect of categorical sorting, they were asked to rate each pair, not in isolation, but after having scanned the remaining pairs, so that each rating should be proportional in relation to the other ratings.

§5.2.1 experiment

Administration: 36 Japanese first year Art students of English Literature participated in the experiment. They are girls from the same district. They were not told that they were going to repeat a similar experiment in the future. The instruction included the requirement to read the whole text three times before answering. Including
the time to deliver the test, the session finished in just 10 minutes. The test consisted of the 7 paired kowasu sentences. Half of the subjects were presented with the scale of 5 $T_g$ and the other half, with the scale of 9 $T_g$. After a three-week interval, the same test was repeated. But the subjects who dealt with the scale of 5 $T_g$ on the previous occasion were presented with the scale of 9 $T_g$ and vice versa. Then 4 weeks after this the same test was repeated, as on the first occasion. Then again 5 weeks later, the same test was repeated and the presentation procedure was the same as on the second occasion (see Fig. 2).

§5.2.2 Result

11 out of 36 failed to complete the task. Of these 11 subjects 8 failed only on the second occasion and 3, on the first occasion. They were omitted from the analysis.

According to the operational definition of a reliable and sensitive scale (see §5.2), (a) group consistency 1, (b) group consistency 2, (c) personal consistency and (d) the sensitiveness of scale are reported below. In calculating statistics, the data are regarded as ordinal. Thus, non-parametric statistics is consistently used: see chapter 7 where the ordinal nature of judgemental data is discussed.

(a) Group Consistency 1

Since memory effect was assumed to be minimal owing to the precautions described above (see §5.2), the number of subjects who were consistent in their reasoning and arrived at the same conclusion on the two occasions is meaningful. The following table gives the number of subjects whose answers were the same on the two occasions and the number of those whose answers were different.

<table>
<thead>
<tr>
<th></th>
<th>Same</th>
<th>Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5T_g$</td>
<td>99</td>
<td>76</td>
</tr>
<tr>
<td>$9T_g$</td>
<td>74</td>
<td>101</td>
</tr>
</tbody>
</table>

$\chi^2 = 7.1438$ ( $0.02 < p < 0.05$ )

Table 23
The scale of $5_{T_g}$ elicited more consistent answers than that of $9_{T_g}$. The difference between them is statistically significant at the level of $0.02 < p < 0.05$.

(b) group consistency 2

Group consistency can be revealed by examining the degrees of association between the two testings. This is achieved in two steps: first, by examining the association of categorical information and second, that of ordinal information. The former was attested by the contingency coefficient $(C)$:

\[
\begin{align*}
C(5_{T_g}) &= 0.75417 \quad \text{upper limit} = 0.89443 \\
C(9_{T_g}) &= 0.83547 \quad \text{upper limit} = 0.94281
\end{align*}
\]

Table 24

$C$ has two limitations: (a) perfect correlation cannot attain unity; (b) since the value of perfect correlation (upper limit) is determined by the number of categories, the value of $C$ is not comparable. However, as is shown above, the two values are fairly equal in relation to the upper limits. These values of $C$ imply that the two scales are very similar, yielding a very high degree of association between the two testings.

It should be noted that $C$ does not take into account ordinal information, but only deals with nominal information. In order to make up for this limitation of $C$, each of the observations is replaced by ranks and Kendall's coefficient of Concordance ($W$) is computed.

<table>
<thead>
<tr>
<th></th>
<th>Test</th>
<th>Re-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5_{T_g}$</td>
<td>0.5849</td>
<td>0.4721</td>
</tr>
<tr>
<td>$9_{T_g}$</td>
<td>0.5125</td>
<td>0.4842</td>
</tr>
</tbody>
</table>

Table 25

Each value of $W$ is equally significant at the level of
Thus, on the basis of C & W we can conclude that the two scales yielded very similar inter-rater correlation.

(c) personal consistency:

Spearman’s rank order correlation (rs) was calculated for individual subjects to measure the association between their two ratings.

(...........see Table 26 in 2.0, Appendices.)

The number of subjects who reached significant levels at $p=0.1$ and $p=0.02$ (two-tailed test) are as follows:

<table>
<thead>
<tr>
<th>Scale</th>
<th>Subjects</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5_{T_g}$</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>$9_{T_g}$</td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>

($r_s \geq 0.714$) ($r_s \geq 0.893$)

$p=0.1$ $p=0.02$ (two-tailed)

Table 27

In terms of personal consistency based upon $r_s$, the scale with $9_{T_g}$ seems slightly better than with $5_{T_g}$.

(d) the sensitiveness of scale:

In both scales, the two extreme categories are explained by the same statements: "identical" and "very different". That is, they are dealing with the same semantic distance and therefore the scale of $9_{T_g}$ is twice as finely divided as $5_{T_g}$. It should, in theory, be twice as sensitive in revealing the differences. Some check upon the degree to which this has been actually achieved may be obtained by calculating some measures to show any significant difference within and between the two scales. An estimate of the extent this is the case might be obtained by comparing the value of $\chi^2$ for every possible combination of test items. Thus, we will get $21 \chi^2$ estimates for each scale. The more effective the scale is, the larger the number of cases will reject the Null Hypothesis at 0.05, 0.01 & even 0.001 probability level. This evidence will support
the effectiveness of the scale.

(...........see Table 28 in 2.0, Appendices.)

The higher the significance level is, the better the scale with 5 $T_g$ appears to be, for revealing significant difference. It can be concluded that the scale with 5 $T_g$ is slightly more effective in our situation.

We can summarize the results as follows:

<table>
<thead>
<tr>
<th>group consistency 1</th>
<th>5$T_g$</th>
<th>9$T_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>group consistency 2</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>personal consistency</td>
<td>✔</td>
<td>✔✔</td>
</tr>
<tr>
<td>effectiveness</td>
<td>✔✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 29

On balance, the scale with 5 $T_g$ appears to be slightly better.

§5.3 the elicitation for MDS

This section gives a brief overview of the methods of eliciting data which will be used as input to the MDS. It concludes that categorical rating (or sorting) appears to be the most economical procedure in terms of time, 'reliability' and "anchoring effects" (see Guilford, 1954, Sandusky, 1974, Parducci, 1974, Anderson, 1974, Jones, 1974, Stevens, 1974.)

§5.3.1 A Brief Overview of MDS

The metric MDS was developed by Torgerson (1958) partly from an earlier sketch by Richardson (1938) as well as some mathematical groundwork done by Young & Householder (1938). The full development of the method is also associated with the psychometric group at Princeton.
University represented by Gulliksen (i.e., Messick, Abelson and Torgerson). By contrast with later developments, the earlier method is often referred to as the "metric" approach to the MDS (cf. Shepard, 1972).

Shepard introduced non-metric approaches to the MDS under the name of Proximity Analysis (1962a & 1962b). Since then, various methods of dealing with diverse types of data have been devised: e.g., TORSCA (Torgerson & Young), Kruskal’s M-D-SCAL, SSA (Guttman & Lingoes), INDSCAL (Carroll & Chang), INDIOSCAL (Carroll & Chang), MDA-OR (Hayashi), MAPCLUS (Arabie & Carroll), etc.: cf. Yoshida 1976; Shepard, 1974; Lingoes, 1977.

A non-technical account of the non-metric MDS rationale based upon Shepard (1962a, 1962b, 1972 & 1974) may be summarized as follows. In performing an elicitation task, the subjects are assumed to be consulting their "mental maps" (in the present case the semantic map). The aim of the MDS is to reconstruct these mental maps with subjective structure, using as input the "proximity" data (i.e., the judgemental data involving similarity). The proximity (similarity) is regarded as conforming to the notion of "distance", although it is a psychological distance. The sense of similarity is between paired stimuli. In the MDS varying degrees of similarity among all possible combinations of pairs are considered to be recoverable in a space with a limited number of dimensions.

The semantic space of words conform to a multi-dimensional space according to Osgood, a (1957), because, apart from discrete semantic distinguishers (e.g., male/female), semantic features can be regarded as continuous variables which can be treated as dimensional axes or continua. In MDS, the subjects' estimates of similarity are assumed to reflect their over-all assessments based upon various dimensional (continuous) features. The MDS aims to restore these criterial dimensions (criterial in the sense that they are used in making similarity judgements) and to represent graphically how input stimuli relate to each other.
present graphically how input stimuli relate to each other
in a reduced space consisting of only criterial dimensions. The dimensions, along with the configuration of points representing each input stimuli, are supposed to be visually interpretable, and to represent some major parameters of judgement among many subjects.

Since the MDS claims to discover some criteria of human judgement, it is often referred to as a discovery or exploratory procedure (e.g., Shepard, 1972). Until Kruskal suggested, in terms of stress values, a method of measuring the extent to which we are imposing a structure with lower dimensions upon the empirical data set, it was more limited as an exploratory procedure (1964a & b). Spence made an important development in this direction (1979). His measure of "stress" which is called "Stress 1" in MINISSA (see below) ascertains that the final configuration of the MDS is not random and that it contains "some genuine underlying structure" (1979:355).

The specific computer programme used in the present research is called MINISSA (Michigan-Israel-Nijmegen-Integrated-Small-Space-Analysis), as in Kelleman (1977 & 1978). Since this is a non-metric MDS, it starts with information about distances rather than with the distances themselves. It requires the ordinal information about every possible combination of paired stimuli. The results obtained from a graph with 2 or more dimensions in which the rank orders of all pairs are faithfully (i.e., depending on stress values) retained. The MDS gives several solutions; a 2-dimensional solution, up to N-1 dimensional solution where N is the number of stimuli. It is up to the user to decide which dimensional solution is appropriate for the data. The clues for this decision are supplied by the numerical information the MDS computes: i.e., the two kinds of stress values and Spence's Stress 1 mentioned above - the latter being available at Edinburough since 1981. This decision-making also depends upon the number of stimuli.

The following hypothetical case will try to give a simpli-
fied account of MINISSA, but for the purpose of illustration, any reference to stress values is omitted. Fig. 3 is a triangular matrix representing similarity judgements among 6 pairs and the number of stimuli is 4. Let us suppose this is the input to the MINISSA.

\[
\begin{array}{ccc}
  A & B & C \\
  B & 3 & 5 & 6 \\
  C & 5 & 4 & 6 \\
  D & 6 & 6 & 2 \\
\end{array}
\quad
\begin{array}{ccc}
  A & B & C \\
  B & 2 & 4 & 5.5 \\
  C & 4 & 3 & 5.5 \\
  D & 5 & 6 & 1 \\
\end{array}
\]

Fig. 3  Fig. 4

The MDS ranks the input data, as in Fig. 4. To speed up the computation, there are two methods, but the main point is that, taking many "iterations", the MDS computes "inter-point" distances, as in Fig. 5

\[
\begin{array}{ccc}
  A & B & C \\
  B & 1.5432 & 1.6039 & 1.7731 \\
  C & 1.5732 & 1.4473 \\
  D & 1.8255 & 1.4473 \\
\end{array}
\quad
\begin{array}{ccc}
  A & B & C \\
  B & 2 & 4 & 5 \\
  C & 4 & 3 & 6 \\
  D & 5 & 6 & 1 \\
\end{array}
\]

Fig. 5  Fig. 6

If we rank the above interpoint distances, we get Fig. 6 One can easily see that Fig. 4 and Fig. 6 are very similar in terms of the rankings of all pairs. This illustrates that the non-metric MDS represents order relations among all the pairs. Although the graph is not presented above, it usually gives a reasonably approximate representation of configuration (i.e., the ways the points representing stimuli are arranged in a space in relation to each other). However, the computational procedure may not be sensitive to the rank order of points along every dimension, as Fillenbaum & Rapoport (1971) point out: the same point is implied by Clark & Clark (1977). It may be unreasonable to expect that all stimulus points fall sensibly in some graded fashion along any single dimension.
As several linguists point out, what Clark & Clark call the "quantificational approach" to meaning tends to represent a structure common to many people and it is not necessarily profound or linguistically significant (Weinreich, 1959, Carroll, 1959, Ullmann, 1962, Clark & Clark, 1977). This "quantificational representation of meaning" (Clark & Clark; 1977: 432) has thus been regarded as yielding a representation of "semantic average" (Weinreich, 1959: 364). However, Clark & Clark state "it has actually been more revealing ... where the investigator's intuitions are of little value" (1977: 435). This statement is particularly appropriate in the case of IL which is concerned with learner-language.

§5.3.2 a brief overview of elicitation of the MDS input:

This brief overview is concerned only with the non-metric MDS. Fillenbaum and Rapoport's experiment (1974) ironically demonstrated that the non-metric MDS requires careful consideration of elicitation and experimental materials. Fillmore took part in their experiment. The materials were "verbs of judging" which Fillmore had analyzed in detail (1971). Even the MDS based upon the linguist's input data did not display his subtle analysis. One reason for this appears to be that the experimenters presented Fillmore with the verbs in isolation; and not even object nouns were provided, or any contextual clues to narrow down the specific meaning of each verb. This indicates that, without giving a specific meaning for each verb, it is not easy even for a linguist to make a proper differentiation among verbs with various meanings. This anecdote shows that experimental materials must be contrived so that the subjects can reveal their linguistic sensitivity. It also indicates the importance of elicitation.

Familiar elicitation methods fall into the following three kinds: (1) "tree method" (Boyd & Livant, 1964, Rapoport, Livant & Boyd, 1966, Rapoport, 1966, 1967 and Fillenbaum & Rapoport, 1971 & 1974); (2) "complete undirected graph" in which the subject ranks all pairs from 1st to the last pair ac-
cording to the 'feel' of similarity. Because of this, some sort of device is usually contrived: e.g., asking the subjects to rate all pairs, an investigator calculates the values for each pair, either by summing all individual ratings (Kellerman, 1977) or by averaging individual scores (Fillenbaum & Rapoport, 1971). Either addition or averaging, these calculations are open to question (see below). The above three methods of elicitation are overviewed in addition to one small experimental inquiry.

(1) the tree method

The subjects write down the two words which are most similar to each other, drawing a line between them and labeling it 1. After this they have 2 options; (i) Among the remaining (N-2) words, pick the word which is the next most similar to either of the two words they have already chosen, and label the connecting line 2. Or, (ii) Start a new tree with a pair which is more similar to each other than to the rest. Until they exhaust the list and all the separate trees are connected into one tree, they can choose either of these options. But they have to label the connecting lines in the order of their recognition of similarity. Thus, every connecting line will be labelled from (1) to (N-1), reflecting the order of recognition. The tree method can be combined with rating according to a given scale (2 samples in Fillenbaum & Rapoport, 1971).

Due to the relative ease of the task, the method can handle a great number of words or sentences at once. However, there is one defect which is revealed in the additive method. The MDS requires similarity (dissimilarity) estimates of all pairs (n sentences means n(n-1)/2 pairs). Whereas, the tree method yields only (n-1) estimates. The remaining (n-1)(n-2)/2 pairs are unknowns. Thus, if there are 10 sentences, the MDS requires 45 estimates, but the tree method gives only 9 estimates. The information about the remaining 36 estimates is missing. To handle these unknowns, Papoport & Fillenbaum adopted the "additive method". For example, suppose the following tree is written
by a subject:

```
A  5  B
  4
C  1  D
  7
E  2  F
  6
G  3  H  8  I  9  J
```

Fig. 7

According to the tree method, the value for the pair A & C is missing. According to the additive method, we simply add the value of the AB pair to the value of the BC pair. Thus, we get the value 9 for the pair A & C. Equally for the value for the pair A & D we add the values of the AB pair, the BC pair and the CD pair. However, if the subject compares A with C as reference point directly, the resulting value may not be the same as (AB + BC = 9).

It may be noted that, in order to get the estimate for the pair A & J, the other intervening pairs are assumed to occur in the order originally specified by the subject; otherwise "addition" does not make much sense. (Besides, it is possible to say that, depending on a specific reference point, the subjective value of a paired comparison changes.)

Fillenbaum & Rapoport define a tree as an "undirected", "connected" graph "without cycles" (1971: chap. 2). It is an undirected graph, since it does not have any reference point for direction. It is a connected graph, since N stimuli are all connected but "without cycle", i.e., there is no crossing. These three features characterize the tree method. In other words the method of addition mentioned above appears to require 3 assumptions:

(a) There is an underlying linear continuum (e.g., by stretching Fig. 7, we get one straight line.)

(b) The rank orders are invariant (e.g., if the order
from A to J changes, we cannot add the values of the intervening pairs.)

(c) The underlying metric is either "city block" or "Euclid." The former means that the subject adopts separable processing, while the latter means that the subject adopts integral processing (as we have briefly seen in chap. 5 (see also chap. 9)). In the former case, the inter-point distance is correctly estimated by addition. But in the latter, the interpoint distance, for instance, A-C in Fig. 7, is not additive (i.e., A-C=√(4^2 + 2^2)). But because of the assumption (a), it is possible to estimate A-C by addition.

Apart from the assumption (b), the other two assumptions are plausible, since the 'feel' of similarity may be based upon the notion of continuum and the cognitively advanced subjects can adopt both separable and integral processing. However, the assumption (b) requires some investigation here. In order to attest this, the following mini-experiment was run. The assumption mean that the initial arrangement of the sentences made by the subject was totally fixed. We had to see whether this is plausible. The procedure takes two steps:

(i) To replicate the tree method.

(ii) Since the tree method adopts a linear and fixed order assumption, one can easily identify the two most dissimilar items in each subject's response. Next, these items are presented one at a time as a standard item with the rest of the set is to be ranked by the subject. If the two rankings show perfect correlation, the invariant assumption is plausible. If not, it means that the subjects change their implicit criteria and that the invariant order assumption is not plausible. (As mentioned above, the tree is not undirected. Thus, the two extreme items can be reference points. This gives us the above experimental rationale: see footnote 10 and p.
I asked 20 native EFL teachers of English to perform the above task using the 9 English sentences of 'break' in appendix 2. Of the 20, only 2 subjects supported the invariant order assumption. The difference between the supporters of the tree method and non-supporters is statistically significant ($\chi^2=12.8$ $p<0.001$). Among the non-supporters, Spearman's rank order correlations are calculated according to the individual subjects. The correlations ranged from 0.4 to 0.96. Thus, this small-scale investigation suggests that the invariant order assumption is not tenable. Rather, depending on the specific pair, the similarity-based order relationship changes. This may be the reason why a tree is defined by Fillenbaum and Rapoport as undirected, which allows for the possibility of different reference points. However, it would appear that for the additive method to be reasonable the intervening pairs must occur in the order specified originally by the subject. On the whole, the only weakness of the tree method is this method of addition. In other respects, the method makes the elicitation task easier than the following "complete undirected graph".

(2) the complete undirected graph

The method is called "undirected", since the relationship between the pairs does not have any specific reference item: i.e., one or the other of a pair is not a standard with which the comparison is made (see chapter 2 in Fillenbaum & Rapoport, 1971). This method is called "complete", since all pairs are joined by some relationship (ibid). The specific instruction for this method is stated by Fillenbaum & Rapoport (1971: 17):

You will be given a randomly arranged list of $M$ pairs of words, where each word is paired with every other word. Go carefully through the list and thoroughly study all the pairs. Then write 1 by the pair which is least (most) similar, 2 by the next least (most) similar pair, 3 by the
next pair ... and so until M for the most (least) similar pair. Work slowly and carefully; this is a difficult task; take your time.

As indicated above, this method is a highly demanding task on the part of the subject. The subject not only must rank all pairs from 1st to last, but must also be reasoning continuously, comparing the large number of remaining pairs. If the task does not actually involve elaborate comparisons, it might mean that the subject is responding in a careless fashion. For this reason, this method of elicitation is not adopted here.

(3) Card Sorting

The subjects are told to sort the sentences written on the cards in terms of similarity of meaning into any number of piles they wish to make. The idea behind this method appears to be that the frequency with which sentences are sorted into the same pile reflects the degree of similarity. Kellerman reports that the subjects spent 5 to 20 minutes and produced 2 to 15 piles. Two relevant points emerge from this observation. The number of piles appears to have something to do with the willingness or unwillingness on the part of the subject. The frequency with which sentences are sorted into the same pile may reflect this attitudinal factor. Further, it does not appear that the simple expedient of sorting into piles gives a clear indication of the degree of similarity on which a particular selection was made.

However, this method is worth modifying. If we combine sorting with rating, this modified method gives us a clear indication of the degree of similarity. This method can be improved even further by adopting an aspect of the tree method. This eclectic method is called here 'categorical rating' and is described below.

§5.3.3 categorical rating (sorting)
This method combines sorting with rating. It also makes use of an aspect of the tree method in that the relationships among categories (i.e., piles) are rated. First, the subjects sort out the sentences in terms of an unspecified similarity criterion into an optional number of groups (subsets), until they feel this is the optimal grouping. Then, several frameworks are given to them as guidance for structuring each subset. The framework depends on the number of sentences in each subset:

![Diagram showing frameworks for 2, 3, 4, 5, and 6 items in a group.](image)

Fig. 8

If they can think of a semantic closeness between items belonging to different subsets, these have to be connected with a line. Then the subjects rate the connecting lines according to a 6-point scale. Finally they are provided with a triangular matrix in which one slot is reserved for each pair of sentences. Thus, by filling each slot with their rating scores, they can check whether any pairwise comparison is missing. In this way, unmotivated subjects are forced to think of the varying degrees of similarity among all the possible pairs of sentences.

§5.3.4 the treatment of individual scores:

So far we have seen that categorical rating is an improvement over the three elicitation methods mentioned previously. The discussion here is restricted to the question of how to elicit the data from individual subjects. These individual scores need to be treated in some way so that group scores (i.e., generalized value over the subjects) can be used as input to MINISSA. The usual method for this is either to add scores or to average them (e.g., addition in Kellerman, 1977 and averaging in Aglin, 1971,
Fillenbaum & Rapoport, 1966 & 1971), MINISSA is a non-metric MDS and is designed to process ordinal information of similarity. In this sense, how the investigator treats the individual scores is crucial. The data treatment by the investigator needs to be congruent with the ordinal nature of the judgement data. Whereas such arithmetic operations as addition, multiplication and division assume the interval nature of data (cf. Guilford, 1956 and Siegel, 1956). If the experimenter also assumes the interval nature of data, there is no need for him to use a non-metric MDS, since the interval data can be analysed by a metric MDS.

However, there seems to be a difference between addition and averaging. In obtaining a value generalized over many scores, the former addition might be less appropriate than the latter method of averaging. For instance, let us suppose that three subjects rate three paired sentences a/b, b/c and c/a as follows:

<table>
<thead>
<tr>
<th></th>
<th>a/b</th>
<th>b/c</th>
<th>a/c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Subject 2</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Subject 3</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

According to addition, the semantic space between the 3 pairs common to these subjects forms an equilateral triangle, which is far from the semantic map of each individual subject - they are all scalene triangles. Thus, summation may result in obscuring and cancelling out individual differences. On the other hand, averaging indicates the central tendency among the individual scores. Averaging is less likely to yield a gloss effect cancelling out individual differences. However, either addition or averaging, these calculations implicitly assume the data to be interval, particularly in the case of ordered series of whole numbers. In this respect, MINISSA being non-metric, these data-treatments might not be satisfactory.

First of all, the experimenter has to recognize the nature
of the information the subjects are capable of supplying for the sort of task we ask. Supposing that, given a 5-point scale, the following three pairs are presented to a subject and that his ratings are as follows:

A. The cup broke. (2)
   He broke his leg.
B. The cup broke. (3)
   The wave broke on the shore.
C. The cup broke. (4)
   The tree broke the monkey's fall.

Note: the subject is supposed to compare the meanings of the verb, break.

From his answer (2, 3 & 4), we can be sure that in terms of the semantic similarity A, B & C pairs are ranked by the subjects in the above order, (but nothing more than this), because there is no sufficient reason to assume that he is aware of the magnitude of the differences between the numerical scores he gave us. We cannot say with confidence that in his mind the difference between 2 (close relationship) and 3 (some relationship) represents exactly the same magnitude as between 4 (slight relationship) and 5 (totally different). The magnitude of the differences, for instance, may be like one of the three hypothetical scales (a), (b), (c), in Fig. 9. That is, we should expect a great deal of individual difference in any subjective scale.

(a) \[ \boxed{\begin{array}{ccccc}
\text{origin} & 1 & 1 & 1 & 1 & 1 \\
2 & 3 & 4 & 5 \\
\end{array}} \]

(b) \[ \boxed{\begin{array}{ccccc}
\text{origin} & 2 & 1 & 1 & 1 & 1 \\
3 & 4 & 5 \\
\end{array}} \]

(c) \[ \boxed{\begin{array}{ccccc}
\text{origin} & 2 & 1 & 1 & 1 & 1 \\
3 & 4 & 5 \\
\end{array}} \]

Fig. 9

In a subjective scale such as the above, numerical re-
sponses are not markers of magnitudes between the units. What is observed consistently in these responses is the ordinal information along each subjective scale.

As we have seen, in translating order relatings into mathematical operations, we cannot, in general, use the usual operations of addition, subtraction, multiplication, division and so on, because they are based on interval units. For example, supposing that all four subjects rate a pair as 4. The scale value for this pair is 4 by the averaging method and 16, by addition. But when the magnitudes between the steps are unknown and found to be unequal as in Fig. 9, these scores of 4 and 16 do not make much sense. Algebraic operations become possible, however, even in the case of an ordinal scale, once the experimenter establishes an operationally defined continuum with an arbitrary origin and magnitude between the steps common to all subjects. This can be accomplished by the Thurstone-Torgerson method.

Thurstone's method (1927a, 1927b & 1954) regards a psychological continuum as having subjective magnitudes. Thus, it is suitable for ordinal data. Thurstone's equations are often referred to as the "Law of Comparative Judgement". Although it is not strictly a law as in physics, it is a well-established method in psycho-physics. Luce (1977) examines Thurstone's model in detail and concludes that both Thurstone's model and his assumptions are tenable even in the light of recent advances in research concerning our central nervous system. In the area of attitudinal measurements there have been a number of applications, Shaw & Wright (1967) report that 64 out of the 175 cases they have examined employed Thurstone's method. There is some similarity between the measurement of attitudes and that of meanings, but Thurstone's method has not been applied to the latter, apart from a few exceptions (e.g., Osgood, 1957, Arnold, 1971, etc.).

As indicated in §5.3.3., this study uses the categorical rating in eliciting input data for the MDS. Due to this
design, the equations used here are based upon Torgerson's law of categorical judgement which follows closely Thurstone's equations (Torgerson, 1958). Torgerson's method is devised to compute category boundaries (see §5.2 and chapter 7) together with scale values for paired stimuli. Thus, the present application of the Thurstone-Torgerson method yields estimates of category boundaries and scale values for paired sentences at the same time. These scale values are ranked and the rankings are used as input for MINISSA (i.e., a version of non-metric MDS). In this way, I can make use of both the MDS results (chap. 9) and the scale values based upon the paired comparison jointly in analysing the translatability judgement data (chap. 10).
Chapter 6  pilot study (3) : free recall experiments

§6.0 purpose

As we have seen in chap.4, the judgemental process of translatability appears to involve some categorization. It was clear that at least the hypothesis of categorization predicted the judgement data better than otherwise. This assumption was made on the basis of the clusters which appeared in the MDS results. We also assumed that the MDS results provided an adequate representation of the NL inter-lexical structure. (This assumption is examined additionally below.)

This pilot study intends to investigate whether or not the categorization of NL items actually occurs in the judgemental process of translatability. This investigation was possible by using a free recall experiment, since free recall reveals that items presented randomly tend to be recalled in conceptually related groups (cf. Tulving, 1972 & Brown, 1976). Thus, free recall is a useful technique to investigate categorization. This study uses two kinds of groups of subjects: an experimental group performs a translatability test first and is asked incidentally to recall the NL items; the control group performs only the free recall task. By comparing the amount of recall in these two groups, one can examine whether or not the task of translatability encourages the subjects to categorize the NL items. An additional interest of this pilot study is to see whether categories (groupings) observed in the free recall tasks can show some correspondence to the clusters produced by the MDS. Further, this study is concerned to discover the nature of groupings which occur most frequently. The specific hypothesis which relates to the present study is given in §6.3. §6.1.gives a brief overview on free recall with special reference to some psychological attributes which influence the amount of recall and groupings. Based upon this overview, experiment (1) is run to see whether the association values for the experimental items show any difference in terms of group scores. The purpose is to check whether or not the following experiment is reasonably controlled. Experiment (2) is concerned with groupings in free recall. The method of assessing groupings closely follows Monk (1976) and is explained in §6.3.2.
In a typical free recall experiment, the subjects are briefly exposed to either a list of randomly arranged items or one item at a time also in a random order. The subjects are subsequently asked to recall the items in any order in which they occur to them. They are not asked to recall items in the order in which they were presented. It has been shown that the subjects tend to recall items together which are conceptually related or associated in certain ways. In the literature, items recalled together are termed "clusterings" or "groupings". Two types of clusterings have been recognized: associative clustering (e.g., Jenkins & Russell, 1952, Jenkins et al., 1958) and categorical or taxonomic clustering (e.g., Bousfield, 1953, Bousfield & Cohen, 1955, Bousfield, et al., 1958, Komori, 1968).

The experiments on associative clustering usually use normative estimates of association (e.g., Kent & Rosanoff's association norms, 1910, Minnesota norms by Russells & Jenkins, 1954, Umemoto's Japanese association norms, 1969a) and show that words with high association strength tend to be clustered in free recall results. Some experiments which also use association norms have demonstrated that clusterings are derived from "inter-item associations" (Desse 1959a, 1959b, Weingartner, 1964, Matthews et al., 1964). According to these experiments, inter-item associations correlate highly with the amount of free recall (e.g., $r=0.88$ in Desse, 1959a). This suggests that clusterings are due to the context of an experiment which contingently creates intra-list associations on the part of the subject. Unless we assume that the formation of association is a part of cognitive activity, clusterings might be the product of some such contingency. It may be possible to regard association as a part of human cognition (cf. Quine, 1973); however, the present study is not committed to this view.

In experiments which reveal categorical clusterings, an experimental list is composed of instances belonging to different categories, such as a category of animals (giraff, camel), musical instruments (flute, clarinet), shellfish (lobster, prawns) etc. The frequency counts of these instances are normally kept constant in these experiments.
Although items are presented to subjects in a random order (one word per three seconds in Bousfield, 1953), the subjects exhibit a tendency to recall items together which belong to the same category (e.g., giraffe jointly with camel). Bousfield and his colleagues interpret categorical clusters as demonstrating that some superordinate concepts stored in cognition provide bases for the subjects recalling items together which are conceptually related. According to this interpretation, free recall is useful in illustrating the presence and the role of conceptual categories in cognition which may be assumed to be a part of linguistic knowledge.

However, Cofer (1965) points out that categorical clusterings might be the outcome of association: i.e., a categorized list facilitates associations which are formed during an experiment and the product of these associations appears to resemble a categorically classified manner of recall. In other words this overt resemblance to linguistic taxonomy among response patterns might be due to what is incidentally initiated by the intra-list associations. Thus, one cannot be confident that categorical clusterings are solely derived from a fragment of linguistic knowledge the subject possesses. To make this cognitive claim about linguistic knowledge involves the complex issues of how memory is structured, how a categorical concept is retrieved, how semantic knowledge is represented in long-term memory, and so forth. For this reason, in the seventies, psychologists increasingly related free recall to the specific cognitive models of retrieval mechanisms and the representation of long-term memory: (e.g., Kintsch, 1974, Rips, 1975, Smith et al, 1974, Meyer, 1970, Quillian, 1969). In this sense, the results of free recall experiments need to be discussed in the light of more controversial recent issues, but these are beyond the scope of this pilot study.

Tulving (1962 & 1964) has introduced a more subject-centred view in that clustering relates to the active strategy of “subjective organization” which the subject adopts in recalling items. According to this hypothesis, it is the subjects who impose organization upon experimental materials. This hypothesis is confirmed by his experiments. Although the experimental lists are not categorically structured nor controlled by association norms, the subjects exhibit some
consistent clustering in their free recall. Tulving interprets this result as indicating that the subject has an ability to organize the items on the list. The same experimental confirmation is also obtained in Bousfield (1964a). According to Mandler's review (1967), these two studies regard subjective organization as occurring in the process of producing recall responses (i.e., in the encoding process). Mandler indicates another possible interpretation in reference to Seibel (1964): i.e., subjective organization may occur while the subject is in the process of comprehending experimental materials; i.e., during the process of decoding input which is later reflected in the subject's output.

The analogical process we have dealt with previously may be related to these processes of categorization, subjective organization and probably association, as association might include a 'feel' of similarity (cf. Richards, 1975). In chap.2 we have seen that the intuitive sense of superordinate concept is derived from the semantic closeness among instances. And Bousfield & his colleague demonstrate experimentally that these superordinate concepts which are stored in cognition at our constant disposal might be the bases for categorical clusterings in free recall. It is always a possibility that subjective organization is based upon the 'feel' of similarity. Thus, it would appear to be plausible to relate the analogical process to the above three processes of association, categorization and subjective organization. In this sense, the present application of free recall experimental technique is relevant to the main objective of the present research. The terminology of subjective organization (SO) is adopted below as a general cover term for clusterings.

Previous experimental studies have established several psychologically relevant attributes which influence SO in free recall. We have briefly looked at the importance of association and category effects in the above. In addition to these, three other factors are also well-known in the literature. The main objective of this pilot study is to show that a free recall experiment is a useful technique in checking our speculative interpretation discussed in the follow-up study of Kellerman's (chapter 4), particularly,
categorization. For this purpose, experimental materials for free recall must be controlled with respect to other factors which are known to bias free recall responses. Otherwise, it is not clear whether the SO we will obtain experimentally is the product of these other biasing factors. The five biasing factors which have been experimentally proved are listed below.

(1) Word association

As we have seen in the above, words with high association values tend to be learnt faster as a part of a list. Word association is normally estimated by a simple basic technique; presenting the subjects with stimulus words, and having an investigator ask them to think of the first and second (even more) words that come to their mind (e.g., Kent & Rosanoff, 1910 and Russell & Jenkins, 1954). These normative estimates of associations are presented in terms of primary and secondary responses (percentage). In Noble (1952), the subjects are asked to produce as many words as possible within 30 seconds. Noble suggests that the average number of associations per 'stimulus word' produced by a large number of subjects in a given time (often 30 seconds or one minute) is an index of the "meaningfulness" of the stimulus word. This definition of meaningfulness (derived from associations) is also adopted by Paivio et al., in their normative estimation of imagery and meaningfulness. Noble (1963) states that meanings psychologists deal with are "intensional" rather than "extensional", "intension" in linguistics being closer to the notion of associations in psychology: a psychological approach intends to convert internal states into externally observable indices. "Stimulus (S) evokes response (R)" is interpreted as experimentally observable semiotics: the meaning of S is R, to put it crudely. Thus, Noble regards the average number of associations per stimulus over many subjects as the index of these "multi-response evocation power". As this definition indicates, the psychological measures of "association" and "meaningfulness" are closely related.
category effects
As we have seen in the above, instances of the same category tend to be recalled together. Category norms are obtained by asking the subjects to write 4 instances in response to a superordinate category (Cohen, Bousfield & Whitmarsh, 1957). In Battig and Montague (1969), the method is to ask the subjects to write down as many instances as possible in 30 seconds. The assessment is combined with the subject's rating of "category potency" (i.e., the subjective assessment of how many members each category has on the codes: (1) none, (4) average and (7) very many).

imagery effect
Images could serve as mediators of recall involving verbal materials. Imagery is defined as "a word's capacity to arouse non-verbal images" (Paivio, et al, 1968:3). Imagery is discussed in terms of two types: Concrete and abstract imagery. These two types are considered in terms of the polarity between concrete and abstract. Thus, although they are distinct types, they fall along the continuum of abstract-concreteness. Concrete imagery is defined "in terms of directness of reference to sense experience" (ibid.). On the other hand "any word that refers to an abstract concept that cannot be experienced by the senses" (1968:5) relates to abstract imagery. In terms of amount of recall and learning, words with concrete imagery are superior to those with abstract imagery (Paivio & Yuille, 1969, Richardson, 1974, Paivio, 1976). This superiority of concrete imagery over abstract imagery is often referred to as the "dominance effect" (e.g., Freedman & Loftus, 1971).

frequency effect
High frequency words have facilitative effects on free recall and they are retained better and more steadily than medium and low frequency words (Postman, 1970). The positive influence of word frequency in free recall has been observed both after a single presentation (Bousfield and Cohen, 1955, Desse, 1960) and in multiple trial free recall (Hall, 1954, Murdock, 1960, Sumby, 1963).
Some high frequency words are also words with high association strength. Because of this overlap, Dimascio (1959) attempted to isolate these two factors of 'frequency' and 'association' using nonsense syllables and showed that in free recall, the frequency of occurrence (i.e., frequency counts see chap.5) is a more influential factor than association (the identical experimental result is also obtained in Imae, 1961). This supports the view that the frequency effect contributes to memory and recall, independently of association. This view is supported by Underwood and Schultz (1960). In terms of the cognitive contribution to learning and recall, the “meaningfulness” derived from Noble’s definition (where the number of associations is regarded as the sole determinant) is not necessarily equivalent to the effect of “the frequency of occurrence”. Incidentally, it is important to bear in mind that the extent in which a learner’s semantic network spreads and consolidates via associations can be an indicator of good learning and steady memory (see §5.1). Since the high frequency of occurrence ascertains up to a point the subject’s greater amount of exposure to a given item, it is tenable to say that the frequency of occurrence is an influencing factor in learning and memory.

(5) cojoint frequency effect
The frequency of co-occurrence of a given category and its instances can affect the amount of learning and recall (Wilkins, 1971). This effect can be explained as the difference in the associative strength between each instance and its superordinate category (Katz and Denny, 1977). As for the normative estimates of conjoint frequency, the method adopted by Battig and Montague (1969) appears to be useful, because it estimates the typicality of an instance relative to a given super-ordinate category (see chap.5). The experiment (1) reported in chap.5 means that we have kept the conjoint frequency effect constant rather than the more objectively assessed “frequency effect”.

196
According to the above summary concerning the major psychological factors, we notice that the five effects above fall roughly into two classes: (1) association & (3) imagery in one class, and (2) category, (4) frequency & (5) conjoint frequency in the other. This rough distinction corresponds to Umemoto's brief remark mentioned in chap.5 (Umemoto, 1969b). While the former relates to the psychological measure of subjective familiarity, the latter is in comparison eligible to the more objective assessment and is called objective familiarity; (2) & (5) relate to the part of linguistic knowledge the subjects may be assumed to possess; and (4) is the frequency count of occurrences in newspapers, journals and books.

In controlling the biasing factors, the following points are considered. The items on the two lists (of kowasu and yaburu sentences) are categorizable in that the subjects might categorize sentences during the experiment according to the object nouns which follow the verbs. The current purpose of the following experiment is to see whether this effect occurs or not, and in this sense the category effect does not need to be controlled. The experiment (1) in chap.5 has ascertained the homogeneous nature of conjoint frequency in the 11 sentences in each list, since they are all 2-frequency items relative to each verb. If a list consists of sentences with different verbs, the frequency counts of these different verbs would need to be kept constant. However, this is not the case in the present situation. Imagery values are not assessed here for the following reasons, words with high imagery tend to correspond to words with high "meaningfulness" values, and therefore have the greater number of associations. In this sense, the investigation into association values can include imagery to a large extent.

Furthermore, Paivio et, al, (1968) regards imagery values as reflected in the ease or difficulty with which mental images involving sensory experiences are evoked by the subjects. The experimenters assume that the subjects are capable of rating this kind of difficulty and ease (1968:4). It will be remembered that the subjects participating in the present experiment could not properly respond to a simpler question in the experiment (1), chap.5 (i.e., "how familiar or rare do you think the sentence is?").

197
Thus, it would not be practical for us to expect them to respond to such a difficult task as rating the difficulty or ease with which images are evoked. Besides a Pylyshyn makes clear (1973), researchers have not reached a consensus with regards to what is imagery (see, also Kosslyn, 1978 & 1980). Thus, the naive subjects (i.e., unaccustomed to experimental tasks of this nature) would formulate various unstable definitions; and it is probable they would confuse their unstable understanding of the meaning of "imagery" with the quick or slow arousal of mental images. For this reason, the experimental assessment of imagery is not taken up here. Since the highly associative words (i.e., highly "meaningful" words) correspond to high imagery words (e.g. Paivio et al.1968), this appears to suggest that the assessment of association values is sufficient for the present purpose.

§6.2.0 experiment (1)

The set of yaburu and kowasu sentences were the same as those chosen by experiment (1) in chap.5. We have seen above that there are five factors which can influence the amount of recall. It is customary to counterbalance these factors prior to a free recall experiment. Pilot study (2) shows that, relative to the groups of subjects who participated, the factor of the conjoint frequency effect is, on the whole, controlled. Since the materials are not artificial but a part of their native language, the degree of meaningfulness in this sense can be assumed as adequate. Besides, the object nouns are learner-generated. What needs to be investigated is whether or not each sentence evokes a similar number of associations. Unless this is ascertained, there is a danger that a higher amount of subjective organization could be the result of intra-list association which may not be related to the semantically based categorization. Large scale investigations concerning association norms and imagery values are available (Umemoto, 1969a and Ogawa & Inamura, 1974). But they deal with individual lexical items alone. It would be possible to sum imagery values corresponding to individual constituent words and to regard the sum as representing the associative or imagery values for the entire sentence (e.g., Kojo, 1979). Words are discrete and grammatically small constituents, but once they
are constituents of a sentence, they are usually regarded as forming a hierarchically arranged tree structure. This structure may not conform to the idea of addition. The meaning of a sentence is always greater than the sum of its parts (Matthew§, 1979:98). It is meanings that produce imagery and associative values. For this reason it appears to be more satisfactory to obtain association values by presenting a sentence as a whole directly to the subjects than by summing the association values of individual lexemes. This small-scale experiment is intended to examine whether a set of experimental lists shows any statistical difference in terms of association values with respect to each group of subjects as a whole. It is preliminary to the following free recall experiments. So a detailed analysis of the content of associations is not undertaken.

§ 6.2.1 procedure

44 first year university students took part in the experiment for the list of 11 kowasu sentences. The same number of students participated for the list of 11 yaburu sentences. For the list of 11 kowasu sentences there were 44 3rd year junior high school students but for the list of 10 yaburu sentences there were 40 3rd year junior high school students. The four kinds of data collection were performed separately. The order of presentation was counterbalanced to avoid fatigue effect. There were 11 sets of item arrangement in which each sentence appeared once in each serial position. Only in the case of 10 yaburu sentences among 3rd year junior high school students were there 10 sets of item-arrangement in which each sentence appeared once in the respective serial position (see p. 307). A small pamphlet of 11 or 10 pages was distributed among the subjects. In association with the sentences presented they were asked to write down as many words as possible, spending 30 seconds per sentence.
§6.2.2 results and discussion

This experiment is designed to discover whether the items on each list produce any significant difference in terms of the total number of associations obtained from each group of subjects. This separate treatment according to the list and the group of subjects is made so as to relate to the data treatment in experiment (2) below. Thus, we tabulate the total frequencies per item separately according to the two lists and the two groups of subjects (see Table 1 to 4). A chi-square test is sufficient, since the data are frequency data and we are only concerned with whether the items on the list are homogeneous in terms of group scores (in the sense of an item's power in eliciting associations, i.e., item "potential" or "potency" as it is called in the literature). Thus, if the null hypothesis is accepted, it means that the items on the list are fairly homogenous relative to the specific group. The degree of "no difference" derived from the acceptance of null hypothesis depends upon the level of probability we will obtain. In this non-standard use of chi-square test, the higher the level of probability is, the better it is in revealing no difference.

Table 1 deals with the kowasu list among 3rd year junior high school students.

(..........................see table 1)

As appendix ( 3.1 ) shows, individual differences can be observed: high association generators are subjects (5), (14), (38) & (42) and low generators, (1), (3) & (35). There is some item difference (e.g., sentences 1 & 2 as opposed to 3, 5 & 6) in terms of average number of associations; e.g., sentences 1 and 6, \( t = 1.60 \) \((0.1 < p < 0.15)\); sentences 2 & 6, \( t = 1.69 \) \((0.05 < p < 0.1)\). However, the present analysis is concerned with whether the list of sentences as a whole produces any significant difference in terms of the total number of associations. One is not concerned with differences in averages, but with the total behaviour obtained from a group of subjects; the response patterns in free recall experiments are again examined in the light of group behaviours. According to the chi-square test, the items on the list show no difference in terms of the total number of associations per item.
Table 2 presents the result for the kowasu list among 1st year university students. The chi-square test indicates that the items on the list reveal no difference again in this group.

(..........................see Table 2)

In terms of the average number of associations, there is a significant difference between items 3 & 6 (t=1.74, 0.1 < p < 0.05). But in terms of the item potentials relative to the kowasu list as a whole and to the group, 11 items do not show any significant difference, as the chi-square indicates.

Table 3 presents the number of associations for the yaburu list among the junior group.

(..........................see Table 3)

In terms of the item potentials relative to the yaburu list as a whole, there is no difference among these 10 items, as shown by the chi-square test.

Table 4 presents the number of associations obtained among the 1st year university students. As before, item potentials relative to the yaburu list show no difference, as indicated by the chi-square test.

(..........................see Table 4)

Thus, the two lists of kowasu and yaburu sentences are shown to be adequate as experimental items for free recall, since, relative to the two groups, each list presents no difference in terms of the number of associations each sentence arouses. In this sense, these items possess highly similar meaningfulness values according to Noble's definition. The experimental lists are controlled in terms of the associative potentials (this experiment and also of the conjoint frequency (experiment 1 in chap.5). By making use these experimental lists, one can claim that clusterings which will be observed in the following free recall experiments are independent of these two biasing factors.
Table 5 compares kowasu with yaburu among the junior students in terms of the mean frequency. The verb 'kowasu' evokes a greater number of associations than the verb 'yaburu'. The difference is significant. The same tendency is observed among the senior groups (see Table 6). The difference between the two verbs in terms of the mean frequency is significant. These results may be related to the psychological 'meaningfulness' and the fact that kowasu is a more general verb than yaburu (see chap.4).

Table 7 compares the junior group with the senior group in respect to the kowasu list. The difference is significant, as the t-test indicates. The number of associations increases roughly by 4 per item, indicating some developmental growth of associations. Table 8 compares the junior group with the senior group regarding the yaburu list. The difference in terms of the average frequency is significant, as is shown by the t-test. The number of associations increases roughly by 5 per item. This indicates some developmental growth in this respect. When we consider the elicitation time of 30 seconds, these increases (4 in kowasu and 5 in yaburu per item) can be regarded as considerable. The difference of age between the two groups is about 4 years. The greater number of association among the senior group could be taken as an indication that the NL semantic network is developed.

(see Tables 5-8, Appendices 3.0)

§6.2.3 summary

§6.1 overviews the five major factors which are known to contribute to the formation of clusterings in free recall. The application of the free recall experiment intends to examine whether some speculation about categorization in chap.4 actually occurs in the judgemental process. Thus, the present application of a free recall experimental technique is relevant, but the relevance is restricted to the categorical clustering. For this reason, except for the categorical effect, the other factors which are known to contribute to the production of clusterings need to be controlled. These other factors are all biasing variables. Among the four biasing factors, the two that are the most crucial under the present experimental conditions are the
association values and the conjoint frequency effect. The justification for the choice of these two is given on p.p. Since we have ascertained by experiment (1) in chap.5 that the experimental materials are constant in terms of the conjoint frequency, this experiment (1) is designed to discover whether the items are reasonably homogeneous in terms of association potentials. The associations were elicited according to the standard method. The frequency data were analyzed separately according to the two lists and the two groups of subjects. The results showed no difference relative to the lists or to the groups of subjects examined. Thus, this experiment (1) confirms the adequate nature of the experimental materials for the following free recall experiments.

§6.3.0 experiment (2)

This experiment has three purposes. (1) Pilot study (1) appears to indicate that the judgemental process of translatability involves the categorization of the NL list: that is, when we assume some categorization, the correlations of the judgement data with the NL improves exceedingly (the same improvement is observed in Kellerman’s study (1977), see chap.5). This experiment intends to examine whether categorization actually occurs in the judgemental process. If the task of translatability judgment reinforces the categorization of the NL list, one can compare the amount of clusterings which are collected under the following two conditions: (a) experimental groups perform the task of translatability first, followed by the incidental free recall (i.e., the subjects are not told to perform a free recall task); (b) control groups perform only a free recall task. If experimental groups produce more clusterings than control groups, it would mean that the judgemental process of translatability reinforces categorization.

(2) The second purpose of this experiment relies on the notion of the size of clusterings. Here only two possibilities are considered. First, it will remembered that the SIM used in Analysis III in chap.5 predicts the judgement data as well as the MDS results. That is, the paired comparison could be the strategy the subject uses. If so, the SO in free recall results should show up as many 2-item clusters.
Second, alternatively, if the subject adopts a strategy of categorizing several items rather than just two (i.e., large chunking), the SO would result in clusterings with more than 2 items. The second purpose of this experiment is elaborated below.

We have seen that in school textbooks published in Japan a learner learns the English word 'break' in association with the NL verb 'kowasu' and a concrete noun as object. We have also seen that before the English 'tear' (yaburu) is introduced, he learns another association of 'break' with 'yaburu'(tear) and this time, the object is an abstract noun 'promise' - at least this is the order in which these words are taught in their schools. The present research simulates this syllabus here and experiments in chap.10. Thus, the subjects in this experiment were initially instructed that the concrete noun mokei (scale miniature) in the kowasu list and the abstract noun hōritsu (law) in the yaburu list can be said in English using 'break'. This instruction establishes the initial association in each group (the two lists are dealt with by two independent groups). Then, they are asked to guess the translatability of the rest of the list.

In the above situation, the subjects are likely to scan the list as a whole and to compare the initial NL stimulus with the rest of the list. Supposing that scanning and comparing are the basic moves, these moves require some specification of the size and extent. The smallest size for scanning and comparison is between two items. Which two items are chosen first by the subjects among all possible pairings? The analogical process might interact in this decision-making. The subjects might choose each specific pair for comparison one by one in the order of the 'feel' of similarity. We will call this strategy 1 which is summarized below. In contrast with this, strategy 2 involves a larger scale of chunking. This strategy assumes that the subjects group the NL items (in terms of the closeness of meaning). This makes the task of judging translatability easier and more economical in that the subjects can exclude semantically remote groups altogether from their consideration. If they adopt this Strategy, free recall would reflect SO consisting of large chunks. One can summarizes strategies 1 and 2 as follows:

204
phase 1: the instruction establishes the initial association between one stimulus word and the English word 'break'.

In the kowasu list, the initial association is formed by the concrete noun 'mokei', and in the yaburu list, by the abstract noun 'hōritsu':

phase 2:

either Strategy 1: the subjects derive their judgement from a sequential paired comparison between the initial stimulus and the other items. Which sequence the subject adopts depends on these steps:
(i) Scan the whole list.
(ii) Compare the stimulus with the other items. Which item is the closest in meaning to the stimulus word?
(iii) Compare one pair at a time in the order of the 'feel' of similarity.

or strategy 2: This strategy assumes that the subjects categorize the NL items in large groupings, using the following three steps:

(i) Scan the whole list.
(ii) Categorize the NL items. Comparing the resultant categories, judge the category first which is the most similar in meaning to the category to which the initial stimulus belongs. Judge the items belonging to this category first, ignoring the semantically remote categories altogether.
(iii) Choose categories in the order of similarity. Proceed the same as before: i.e., to judge the translatability of the items belonging to the same category.

phase 3: The possible implication of strategy 1 is that there will be more 2-item clusterings than larger clusterings. Strategy 2 on the other hand implies the reverse to strategy 1 in terms of the size of clusterings in free recall.
(3) The third purpose of this pilot experiment relates marginally to the dominance effect mentioned on p. 175 point(3). As we see in chap.10, this dominance effect relates to an IL aim of discovering some cognitive (and therefore language-independent) characteristics in second language acquisition. The initial stimulus in the kowasu list is a sentence with the concrete noun ‘mokei’ and in the yaburu list, a sentence with the abstract noun ‘hōritsu’. According to the dominance effect, the stimulus in the kowasu list should be more facilitative. This superiority of concrete items over abstract items is going to be examined later in chap.10 with respect to the dominance effect upon vocabularily learning. In this pilot experiment, we take some interest in whether the difference in the initial stimuli can cause some variations in terms of the amount of clusterings. The result is discussed in relation to the clusters depicted by MDS. We will look at whether the MDS clusters correspond to the subjective organization in free recall.

§ 6.3.1 procedure

§ 6.3.1.1 the method of presentation

Instead of presenting one sentence at a time the whole list presentation is adopted, since Winograd et al., (1971) show that for categorized lists the method of complete presentation does not differ significantly from the typical single item presentation. Furthermore, the constant order of presentation over trials does not differ significantly from the varied presentation over trials. They argue that while the constant order of presentation facilitates recall for the unrelated words, this potential advantage of a constant order becomes unimportant in the case of a categorized list of words, since “the subjects have the opportunity to organize recall on the basis of conceptual categories” (1971: 225). The present pilot study is concerned with obtaining generalizable results for a group of subjects with regard to the question of whether the task of translatability judgements involves some categorization. The experimental validation of constant order vs varied order and a whole list presentation vs a single item presentation is to be pursued in theory but it is extraneous to the present purpose. For this reason merely on the strength of Winograd et al,
the feasibility of using many subjects at the same time the whole list of 10/11 sentences is adopted. The order of presentation for each individual is the same, but it varies across individuals so that the drawback of the constant presentation can be cancelled out in terms of the group behaviours as a whole. Since the order varies across individuals, primacy and recency effects on memory can be counter-balanced as well. Each item is presented so that it occupies each of the 11 serial positions:

( Supposing that random numbers are renumbered 1 to 11 as in the first row. We rearrange the renumbered items as in the other rows. )

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>
There were two age groups: 3rd year high school students and 1st year university students. Each age group has control and experimental groups. The former group does not perform the task of translatability judgements. For the list of kowasu sentences 22 students took part in the experiments. For the list of yaburu sentences there were 22 students in the university group but 20 students in the junior high school group. The list was written on the pamphlet of 6 pages for the control group; the first page was a blank sheet followed by the list. When the subjects turned the page, there was another blank sheet and so on. The answer sheet consisted of six blank sheets. For the experimental group the sentences of the list was numbered. The stimulus item was circled to indicate that it is the translatable item. In other respects the pamphlet is the same as in the control group. The subjects were told to write down the number of the sentences which, they feel, can be said in English using 'break'. The answer sheets were the same as those for the control group.

The three trials were repeated. Between trials they performed the following distractor tasks: mental arithmetics (3 minutes between trials 1 & 2 and 2 & 3); listing the names of western dishes (2 minutes between trials 1 & 2) and listing the names of prefectures (2 minutes between trials 2 & 3).

The experimental group wrote down their judgements of translatability with the numbers which headed each sentence (2 minutes). They were told to turn the page of both pamphlet and answer sheet and asked to recall the sentences in any order (2 minutes). Then, they did mental arithmetic of addition, while two-digit numbers were orally presented. (12 additions in all and about 15 seconds per addition). Then, they were asked to write down the names of western dishes between trials 1 & 2 and the names of prefecture between trials 2 & 3. Including the delivery of the pamphlets, the experiments took about 30 minutes (22 minutes for the actual task for the experimental group). Among the university students the experiment was administered toward the end of the normal lesson (90 minutes per lesson). Among the junior high school students the experiment was administered in the
presence of their teacher. The control group followed the same procedure as the experimenter group except for the task of translatability judgement. They were told to recall the sentences in any order.

§ 6.3.2 the method of counting clusterings

The method closely follows the hierarchical grouping analysis (HGA) developed by Monk (1976). The method is specially designed to analyze SO in free recall. According to Monk, HGA is a development of the familiar methods proposed by Bousfield (1953), Bousfield (1965), Tulving (1962) and so forth (see Sheull, 1969). Monk points out the following weaknesses in the previous methods. First, only two trials are considered and if a group of items occur in the same order on both trials, these items are counted as one subjective organization. The criterion of the same order over the two consecutive trials (inter-trial repetition in Bousfield’s terminology) cannot avoid the danger of chance results. Second, the size of subjective organization may be larger than a clustering consisting of two items.

Referring to Pellegrino (1971), Monk indicates the possibility of a larger cluster in subjective organization. In the following example, the four items 'cup, glass, china, mug' clearly constitute one conceptual unity, i.e., a category of dishes. When one reads through the items recalled, one can see intuitively that the category of dishes is remembered by the subject, followed by the category of animals.

Example 1

Trial 1: cup, glass, china, mug, giraph, dog
Trial 2: glass, cup, china, mug, dog, camel, elephant

However, according to the previous criterion of inter-trial repetition, the obvious conceptual unity observed in the four items is not accommodated in the analysis. The order of 'mug' and 'china' being constant over the two trials, only these items are regarded as forming a cluster. This is not satisfactory. Although the four items do not appear in the same order, they appear nevertheless in a variable order. If we incorporate this variable order into an analytical framework, the obvious categorical clustering consisting of
the four items can be accommodated. Supporting Pellegrino's view, Monk also proposes the method which includes constant order, reverse order and variable order (see below) so that subjective organization involving more than two items can be taken into account. Third, Monk recognizes the great potential in "cluster analysis": even a hierarchically arranged tree structure may be required in order to recaputune subjective organization.

In the present experiment three trials are performed. Since Monk's criterion of consistency with respect to the constant (reverse) and variable order is defined over three trials, his definition of three kinds of internal orderings are adopted here:

1 variable order (VO)
The items "occur together but in different output orders in the set of trials" (1976:4)

2 constant order (CO)
The items "occur together and in the same order on each trial" (ibid.)

3 serial order (SO)
The items "occur together in the order stated and its reverse" (ibid.) The two-item SO is arbitrarily classified as 2-item VO (ibid.).

The above definitions are illustrated, using the data obtained by the present experiment.

Example 1 from the data

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>7</th>
<th>3</th>
<th>4</th>
<th>1</th>
<th>9</th>
<th>10</th>
<th>5</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 2</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Trial 3</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>11</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

Items 3 & 4 occur in the same order over the three trials. 3 & 4 is an instance of 2-item CO, and it satisfies the definition of consistency over three trials. On the other hand, items 9 & 10 occur in the same order in trials 1 & 3, but the order is reversed in trial 3. It satisfies the definition of consistency and illustrates 2-item VO.
Example 2

Trial 1 1 7 10 9 2 5 3 8
Trial 2 10 7 1 11 2 9 5 6 3 8
Trial 3 11 2 9 10 7 1 5 4 6 3 8

The order of items 1, 7 & 10 is reversed in trial 2 and the original order is restored again in trial 3. Thus, items 1, 7 & 10 are an instance of 3-item SO. Items 9 & 2 is an instance of 2-item VO. Items 3 & 8 are an instance of 2-item CO. It is to be noted that items 9, 2 & 5 occur in trials 1 & 2, but in trial 3, the cluster is broken by the intervening items 10, 7 & 1. When the three kinds of order (CO, VO & SO) is broken, it does not satisfy the criterion.

Example 3

Trial 1 1 7 9 2 5 6
Trial 2 7 1 10 9 2 11 5 3
Trial 3 1 10 9 11 2 5 3 8 6

In the above example 3, the group of 9 & 2 is broken by the item 11 in trial 3. Thus, this example does not possess any group.

In some cases CO, VO and SO are recognized among groups. The definitions of CO, VO and SO are the same as the above. They must satisfy the consistency over three trials.

Example 4

Trial 1 1 10 7 11 2 3 4 5
Trial 2 10 1 7 2 11 3 4 5
Trial 3 1 7 10 2 11 3 4 5 6

Example 5

Trial 1 4 6 7 1 10 9
Trial 2 7 1 4 6 10 9 2 8 3 5
Trial 3 7 1 6 3 10 9 2 8 5 4

In example 4, items '1, 10 & 7' illustrate 3-item VO, items '11 & 2', 2-item VO and items '3, 4 & 5', 3-item CO. These 3-item VO, 2-item VO and 3-item CO constitute CO. In example 5 items 7 & 1 and items 10 & 9 are both 2-item CO. Although these 2-item COs appear in the same order over the three trials, the consistency is broken by the intervening items 4 & 6 in trial 2 and again by items 6 & 3 in trial 3. Thus,
according to the above definition of consistency there is no CO between these 2-item COs.

§ 6.3.3 results and discussion

As we have seen in 6.3.0, we have the three points of interest:

(1) Whether or not the experimental groups yield more subjective organization. If so, the task of translatability judgement encourages the subjects to categorize the NL items (see pp.263-4).

(2) Which is the dominant strategy, strategy 1 (paired comparison) or strategy 2 (involving large chunkings) (see pp.267-6).

(3) Whether or not the dominance effect relates to this experiment. (3) Additionally, we are interested to see whether the clusters depicted by the MDS correspond to subjective organization obtained in this free recall experiment (see pp.215-6).

We will look at the result, according to the above three purposes.

(1) Table 9 summarizes the contrast between the experimental and control groups in terms of the number of clusterings obtained. In all four conditions, the experimental groups produce more clusterings. The differences between the experimental and control groups are significant (see the four chi-square tests in the table). The mean differences are also significant in the three cases, but the t-test is not applicable to the junior group involving yaburu, since the variances are not equivalent in the experimental and control groups.

(...................see Table 9)

Thus, the above results suggest that in making translatability judgements the subjects categorize the NL lists and that these categorizations are reflected in the greater number of clusterings among the experimental groups.

However, according to the above definition of consistency
with respect to CO, SO and VO, some subjects did not produce any grouping, while one subject produced as many as 9 groupings (see 3.0 Appendices). Thus, there are undeniable individual differences in the amount of subjective organization in free recall. As to the interpretation of the above chi-square tests we draw a tentative and conditional conclusion that in terms of group behaviours there is a significant difference between the two experimental conditions (with or without a translatability test), although it is likely that an individual's reliance on categorization varies. On balance it could be said that the task of translatability judgements involves some categorization of the NL list on the part of the subjects.

(2) Table 10 tabulates the number of 2-item groupings in contrast with groupings which involve more than 3 items. The purpose of this tabulation is to show that 2-item groupings are the more preferred form of subjective organization.

(..................see Table 10)

2 out of 4 experimental conditions show that there are considerably more 2-item groupings than larger groupings; the differences are statistically significant. Although the other two experimental conditions do not reach a significant level, the 2-item groupings are favoured by the subjects (see Table 10). This suggests that Strategy 1, the aspect of paired comparison, (and the sequential aspect of Strategy 1 is discussed in (c) below) is persistently used in all these experimental conditions. In this sense, the use of Thurstone-Torgerson method of measurement (SIM) in chap. 10 is justified by this free recall experiment since SIM is based upon the paired comparison. According to our hypothesis mentioned in 6.3.0, point (2), if the subjects adopt a strategy of paired comparison in making a translatability judgement, the outcome of this strategy would show up in free recall as the predominant occurrence of 2-item groupings. This prediction is supported by the data obtained.

Strategy 2 involving large groupings also appears to be in operation (see Table 10, the second column). This is not necessarily contradictory to the experimental confirmation of Strategy 1 mentioned above. It would be more real to the actual process for us to postulate the judgemental process
as Strategy 1 being predominant but modulated by Strategy 2, rather than for us to determine one or the other in a clear-cut manner. This seems to indicate that in analysing the data (chap.10) one can make use of the MDS results jointly with the pairwise measurement derived from the Thurstone-Torgerson method, because the MDS representations are useful in showing us the presence of larger clusters, while the Thurstone-Torgerson method accords with the predominant 2-item groupings. This use of the MDS result is also supported by point (c) below which illustrates that the large clusterings observed in free recall correspond to the clusterings depicted by MDS.

The number of large clusterings shown by the four conditions is related to the analysis of “data processings” in chap.9. The large clustering in the free recall data are always composed of constituent subclusters (see, 3.2, Appendices). Thus, the number of large clusterings can be regarded as an index indicating how well-structured the NL interlexical structure is. As the third column of Table 10 shows, according to this index, the kowasu list which is dealt with by the university students is the most well-defined of all and the yaburu list by the junior high school students is the least structured. The other two experimental conditions come equally midway in the course of development (and it is interesting to notice that kowasu (break) is more general verb and is acquired by L1 children earlier than yaburu (tear), according to several Japanese mothers and British mothers):

<table>
<thead>
<tr>
<th>univ.</th>
<th>Yaburu</th>
<th>Kowasu</th>
</tr>
</thead>
<tbody>
<tr>
<td>least structured</td>
<td></td>
<td>most structured</td>
</tr>
<tr>
<td>junior</td>
<td>Yaburu</td>
<td>Kowasu</td>
</tr>
</tbody>
</table>

The above pattern of development in terms of the NL interlexical structure agrees with Shepp’s developmental analysis of “data processing” (from integral to separable processing) discussed in chap.4.

(3) & (3)' The point at issue here is to see (a) whether the dominance effect (see point (3) p.195 & p.204 in this chapter) is observed or not; (b) whether the clusterings observed in the free recall experiments correspond to the clusters produced by MDS (if the correspondences are fair,
one can prove some psychological plausibility with the MDS technique); (c) we have ascertained the predominant use of paired comparison in (2) p 2/3, which is one half of Strategy 1, but this strategy 1 has another counterpart, that is, sequential scanning in the order of similarity (see pp. 204-5) -- this aspect is examined below.

In Tables 11-18 tables with odd numbers present the tabulation of the four kinds of recall data according to the clusters depicted by the MDS (see §3.4 & 6.1, Appendices). They are called here discrete groupings, because they correspond to the apparently 'discrete' (i.e., without overlaps) clusters produced by the MDS. Whereas lower tables with even numbers present the tabulation of the clusterings recalled which involve two or more discrete MDS clusters. These clusterings are called here 'cross-cluster groupings', since the name suggests that the subject is cross-referencing the NL items across categories, trying to arrive at his decision-making. These latter tables with even numberings are useful to reveal how far the sequential scanning extends in the judgemental process. There are considerable numbers of cross-cluster groupings, particularly in the case of yaburu.

Kowasu  (university)

<table>
<thead>
<tr>
<th>Concrete object nouns</th>
<th>Middle(M)</th>
<th>Abstract(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C₁ : (1, 7, 10)</td>
<td>C₂ : (2, 9, 11)</td>
<td>(5, 6)</td>
</tr>
<tr>
<td>No.</td>
<td>19</td>
<td>16</td>
</tr>
</tbody>
</table>

Table 11  Discrete groupings

<table>
<thead>
<tr>
<th></th>
<th>C₁/C₂</th>
<th>C₁/C₂/M</th>
<th>C₂/M</th>
<th>C₁/A</th>
<th>C₂/A</th>
<th>M/A</th>
<th>C₁/C₂/A</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 12  Cross-cluster groupings

215
Kowasu  (junior high school)

<table>
<thead>
<tr>
<th></th>
<th>Concrete object nouns</th>
<th>Middle(M)</th>
<th>Abstract(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS</td>
<td>$C_1: (1, 7, 10)$</td>
<td>$C_2: (2, 9, 11)$</td>
<td>$(5, 6)$</td>
</tr>
<tr>
<td>No.</td>
<td>13</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 13  Discrete groupings

<table>
<thead>
<tr>
<th></th>
<th>$C_1/C_2$</th>
<th>$C_1/M$</th>
<th>$C_2/M$</th>
<th>M/A</th>
<th>$C_1/C_2/M/A$ (All)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 14  Cross-cluster groupings

The discrete groupings in Tables 11 & 13 represent the number of free recall clusterings corresponding to the clusters in the MDS results. The cross-cluster groupings in Tables 12 & 14 represent the number of free recall clusterings which do not correspond to the clusters in the MDS.

The differences in frequency between the discrete and cross-cluster groups are significant: $\chi^2 = 5.9754$ ($0.01 < p < 0.02$) in the university group; $\chi^2 = 9.0$ ($0.001 < p < 0.01$) in the junior group. Thus, in the case of the kowasu sentences, the free recall clusterings correspond significantly to the MDS clusters and, by implication, the result confirms some psychological plausibility of the MDS technique.

Item 1 is the initial stimulus. Grouping called $C_1$ (to which item 1 belongs) is often revealed in free recall clusterings as combined with the other MDS clusters. By examining the way in which $C_1$ is combined with other groups one can have a fair picture of how the subjects were cross-referencing (see, Table 12 & 14).

It is interesting to notice that the university group shows the more varied cross-clusterings. This may relate to the fact that the cognitively mature subjects can cope with the task more efficiently. The older subjects are likely to be able to examine the various clusters simultaneously (either in parallel or sequentially). It is arguable whether the ability of cross-reference requires parallel or sequential
processing. This may indicate one flaw in Strategy 1. Strategy 1 assumes that 2-item groupings are based upon sequential scanning. But 2-item groupings are not necessarily the outcome of sequential scanning. Even when the subjects compare large NL clusters in parallel, as long as the paired comparison is the basic strategy, the outcome will be revealed as 2-item groupings in free recall results. In this sense, our speculation about the form of scanning needs some further investigation, but this is a future topic of research. The paired comparison simply indicates the comparison of any pair at a time. The paired comparison does not presume whether two items belong to the same cluster or not. Thus, the large amount of 2-item groupings obtained in the data confirms that the subjects adopt a strategy of paired comparison in their judgemental process of translatability, but the data cannot determine whether a form of scanning is sequential or in parallel. However, we can indicate that the latter would be more effective, particularly when the subjects compare large clusters simultaneously.

Since the 2-item clusterings are predominant according to the result in (2), this aspect of Strategy 1 is supported by the data. But according to another result in (2), larger clusterings are also in presence. And the result discussed in (3) & (3)' shows that the large clusterings correspond to the MDS clusters. Thus, on balance, as in the discussion presented in (2) above, Strategy 1 combined with Strategy 2 would be closer to the actual judgemental processes. Thus, in the following analysis of the kowasu data, chap.10, both the MDS results along with the Thurstone-Torgerson method will be used in the case of the kowasu sentences.

Regarding the dominance effect on the free recall clusterings, the effect at first glance appears to exist in the kowasu data. Excluding the Middle category, the university group shows a significantly larger amount of clustering consisting of concrete items as opposed to the clusterings consisting of abstract items: $\chi^2 = 30.8571 \ (p < 0.001)$. A significant difference is also obtained among the junior subjects: $\chi^2 = 15.2439 \ (p < 0.001)$. However, when we compare this affirmative evidence for the dominance effect with the case of yaburu, the evidence can be seen as the direct outcome of the different input. Because the initial association was
established by the concrete noun ‘mokei’ (scale model) in the case of kowasu, we have obtained more clusterings consisting of concrete items. But, in the case of yaburu the initial association being established by the abstract noun ‘horitsu’ (law), we have obtained more abstract clusterings (see, Tables 15 & 17). Thus, the dominance effect ought to be examined, not by the immediate free recall situation, but in relation to the difference of input.

Tables 15 & 16 present the results for the yaburu sentences among the university students, and Tables 17 & 18, among the junior high school students. The MDS result among the university students is from the 2-dimensional solution and among the junior high school students it is obtained from the 3-dimensional solution, since the 2-dimensional solution only produced two clusters.

Yaburu (university)

<table>
<thead>
<tr>
<th>Concrete(C)</th>
<th>Middle(M)</th>
<th>Abstract(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 15 Discrete groupings

<table>
<thead>
<tr>
<th></th>
<th>C/A₂</th>
<th>A₂/A₃</th>
<th>A₂/M</th>
<th>A₁/C</th>
<th>C/M</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁/A₂</td>
<td>16</td>
<td>16</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>A₁/M/C</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A₂/A₃/M/C</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 16 Cross-cluster groupings

218
Yaburu (junior high school)

<table>
<thead>
<tr>
<th>Concrete (C)</th>
<th>Middle (M)</th>
<th>Abstract (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS</td>
<td>No.</td>
<td></td>
</tr>
<tr>
<td>(5, 8)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>A₁ : (2, 6)</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A₂ : (3, 4, 9, 10)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>A₃ : (7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 17 Discrete groupings

<table>
<thead>
<tr>
<th>A₁ /C</th>
<th>A₁ /A₂</th>
<th>A₁ /M</th>
<th>A₁ /A₃ /M</th>
<th>A₁ /A₂ /M</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A₁ /A₂ /M /C</th>
<th>A₂ /C</th>
<th>A₂ /A₃</th>
<th>A₂ /M /C</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 18 Cross-cluster groupings

One common characteristic among these two groups is that there are more cross-cluster groupings here than in the case of the kowasu sentences. In the university group there are 18 cross-cluster groupings, as opposed to 8 in the case of the kowasu list; in the junior group 10 groupings in contrast with only 5 in the kowasu list. This indicates that the initial input of the abstract item made the subjects cross-reference the other categories. It must be born in mind that this difference in input is the only obvious difference between the two experimental conditions.

In the case of the Yaburu list, the judgemental process apparently involved more cross-references and sequential scanning. This could be the reason why the MDS clusters dealing with Yaburu did not accord with the free recall clusterings as much as those with Kowasu. Among the junior students the MDS clusters agree with the free recall clusterings better than among the university students (see Tables 15 & 17). This might be related to the speculation (made on p.67) about the scanning speed: the elder subjects being capable of scanning the list more efficiently, the MDS clusters (which may be assumed to offer a reasonable approximation to the internal representation of the NL inter-lexical structure) were not revealed in the free recall
clusterings; rather, the outcome of the scanning were revealed in the free recall clusterings. On the other hand, in the case of the junior subjects, they would scan and cross-reference the items on the list much more slowly. Thus, it is likely that the MDS clusters were revealed intact without the internal NL interlexical structure being distorted by their scanning. This may account for the reason why the free recall clusterings present better correspondences with the MDS clusters than in the case of the university students. If the contribution of different scanning speeds between the two groups were a reasonable interpretation, there is nothing much for us to doubt some psychological plausibility in the MDS technique. Thus, as in the case of the kowasu sentences, one can draw a conclusion that Strategy 1 along with strategy 2 would be the more real to the judgemental process on the condition that Strategy 1 has more weight in the assessment. This means that, although the MDS results have to be consulted, the Thurstone-Torgerson method offers a more effective measurement.

§6.3.4 summary

1 The task of translatability judgements increases the amount of subjective organization. This is demonstrated by the difference between experimental groups and control groups in terms of the total number of groupings.

2 There were more 2-item groupings than groupings involving more than 3 items. This supports the use of Thurstone's measure of paired comparison in the present study.

3 The groupings in the free recall experiments analyzed by Monk's method exhibit fair correspondences with the clusters represented by the multi-dimensional scaling.

4 The dominance effect was observed in the data; the sentences with concrete object nouns were recalled more often than those with abstract object nouns.

5 The analysis of the data obtained from the incidental free recall (which was preceded by the translatability judgement) revealed that the judgemental process
involved both Strategies 1 and 2 (see 6.3.0). The hypothesis of Strategy 1 is that the subjects derive their judgement from the sequential paired comparison. But this Strategy involves (i) the scanning of whole list and (ii) the comparison with the other pairs, to choose the item closest to the 'sure' item. The former subprocess of scanning may help the development of large chunkings (and could involve some parallel processing). The latter subprocess of comparison will bring about the smaller chunkings. Thus, although Strategy 1 is essentially based upon the sequential paired comparison, this Strategy will help the growth of large and small chunkings.

Strategy 2 was assumed to start with large chunkings and end in the paired comparison. Since the data showed the heavy reliance upon the 2-item groupings, we have concluded that Strategy 1 were combined with Strategy 2 in the judgemental process of translatability. It is interesting to find out whether the translatability data support the analysis above.
Chapter 7 pilot study (4): experiments concerning the application of the Thurstone-Torgerson method

§ 7.0 Purpose

The Thurstone-Torgerson method (SIM) is used in this research to estimate the judgemental limit as well as the 'feel' of similarity between paired items. Two experiments are conducted to examine whether the present application is appropriate for this purpose. This investigation is important for three reasons. First, while MDS cannot compute the judgemental limit, SIM can evaluate the judgemental limit. Second, pilot experiments (1) and (3) show that the paired comparison is the basic strategy the subjects have adopted. Third, the estimation of judgemental limit is important to understand the judgemental limit of translatability (see chap.3).

(1) We have seen in analysis III that when the subjects are assumed to be adopting integral processing, the Thurstone-Torgerson method predicted the analogical process as accurately as the MDS method. Thus, SIM is as important a technique as MDS. Table 1 summarizes the relationships between processings and these two methods. The first three processings have been discussed in chap.4 and the latter two are discussed in chap. 9. The purpose of Table 1 is to indicate that SIM is as efficient as MDS in relating the derived measures of similarity judgement to processings.

<table>
<thead>
<tr>
<th>processing</th>
<th>MDS metrics</th>
<th>SIM</th>
<th>MDS metrics</th>
<th>SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) integral</td>
<td>+ Euclid</td>
<td>+</td>
<td>* Euclid</td>
<td>+</td>
</tr>
<tr>
<td>well-structured</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>poorly structured</td>
<td>+ city-block</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) separable</td>
<td>+ city-block</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) asymmetric separable</td>
<td>+ city-block</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d) dominant</td>
<td>+ supremum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e) categorical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1

As Table 1 indicates, SIM is, in one essential way, distinct
from MDS with respect to the present application of these methods. That is, when the NL knowledge is well-structured, MDS is more useful (see §4.4.3). Whereas, if the NL knowledge is poorly structured, SIM is more useful. This difference is best illustrated in reference to the respective assumptions made in these methods. MDS assumes that subjects are capable of responding to the constituent dimensions which are inherent in the input stimuli (see chap. 9). If this assumption is not plausible with regards to the specific subjects participating in a given experiment, dimensional representation by MDS is not realistic or meaningful, since the application of MDS in such a situation means that the experimenter is forcing upon the elicited data a dimensional structure which may not actually exist: the same point is indicated by Shepard, 1974.

One merit of SIM is that it does not require a well-structured input.

The degree of NL development is clearly relative to the subjects participating in a specific experiment. SIM, the way the method is used here, is suitable not only when the NL interlexical structure is well-developed but also when the fragment of NL happens to be poorly structured. In the case of a well-structured NL, MDS is apparently a superior technique which provides us with more numerical information. But SIM, being mathematically simpler, has its unique use, particularly when part of a learner's mental grammar under investigation is not well-defined. Because of this characteristic, we have recognized the usefulness of the method as being more suitable than MDS for explaining the beginner's IL judgemental process (see chapt. 4).

(2) we have seen in chap. 6 that the majority of clusterings in the free recall data are two-item based, suggesting that paired comparison (on which the Thurstone-Torgerson method is based) is the predominant operation which the subjects have adopted during the judgemental process of translatability. It will be remembered that this tendency was more marked in the experimental groups.
than in the control groups. Thus, the foregoing experiments endorse the important role of paired comparison in the judgemental process. The notion of paired comparison being relevant in the present research, the adequacy of the method of measurement for this simple analogical process needs to be tested by an independent experiment.

The Thurstone-Torgerson method is based upon the idea of paired comparison, but it is essentially a technique whose cognitive relevance is not accurate (in every case). The method may be satisfactory, but in some sense it is not flawless. Thurstone and Torgerson call their equations "the law of comparative judgement" and "the law of categorical judgement" respectively. But, as indicated by Yoshida(1976), these psychophysical laws are not the same as natural laws. For instance, the method uses an arbitrary origin (i.e., a zero point) which is distinct from an absolute origin( see Torgerson, 1958). In addition, rather than solving the unkowns in the equations mathematically prior to actual calculation, the method makes a few simplifications whose psychological adequacy in the present application remains to be seen—cf. the impressive demonstration in the area of sound perception by Luce(1977) cited (see footnote 7 in chap.5 ). Conversely speaking, it would appear that because of the problem of arbitrary origin and some computational simplification, the scale values derived from the method must be examined through some independent experiments.

The main purpose of the following experiments is not to examine the adequacy of the equation itself, but to examine whether the present application of the method is justifiable in view of the material used and the subjects participating. In addition to the necessity of this experimental examination, it is important to bear in mind that the method makes use of normal distribution function extensively. Moreover, following Thurstone(1932), the present application involves the further assumption of the underlying linearity of similarity continuum. Thus, the following experimental
examination is indispensable. Whatever justification we may have is relative to the subjects and the material used and may not be generalized. Any subsequent user of this method would require further experiments to check whether the various assumptions will accord with the nature of new data and subjects.

§ 7.1 Overview

Since both Thurstone's method (1927a, 1927b & 1932) and Torgerson's method (1958) are familiar, a brief summarization will suffice for this overview. Torgerson’s method closely follows Thurstone’s, applying it to the categorical judgement which elicits responses according to some ordered categories (see Torgerson, 1958). The present application is another example of categorical judgement. The method is explained below mainly in reference to Thurstone’s, but the source of this summary comes from Torgerson (1958). This method is usually illustrated by psychophysical responses, but here some examples are drawn from the present experiment.

We take a series of NL noun-verb combinations as given. Since we deal with two polysemous verbs, it is reasonable to regard the respective lists of NL items as a series. To this series, the subject responds differentially with respect to a given attribute or a bundle of attributes. Psychophysical experiments tend to deal with the former case of a single attribute; and other applications, such as measurements of attitudes, typically deal with the latter case of a bundle of attributes. The Thurstonian model aims to locate a series of discrete stimuli on a subjective continuum. Here, we wish to locate a series of all the possible pairs of the NL sentences so that we have a representative scale of values for all these paired sentences. However, in the present application, we are not concerned with constructing a scale on which each NL sentence is scaled, but rather with a similarity continuum along which each pair of all the possible combinations can be located, since our central concern is to quantify
the 'feel' of similarity between two items. The location of each pair along the continuum which is derived from the Thurstonian model is called "scale value" (SV), following Thurstone's terminology. As we see below, SV is nearly similar to our usual notion of "average" either over many subjects or over many responses made by a single subject. This is mainly because the equation is derived from the normal distribution function (see Torgerson, 1958, chap. 8).

Both Thurstone's and Torgerson's methods involve two postulates. Each stimulus is assumed to give rise to a "discriminal process", due to the given property or bundle of properties mentioned above. Each discriminatory process is postulated as conforming to a normal distribution function with a mean (which is designated by Thurstone and Torgerson as "scale values" (SV)) and a standard deviation (i.e., "discriminal dispersion" in their terminology). In this sense, the Thurstone-Torgerson method is familiar to us. However, there is a difficulty here.

The difficulty is that the subject cannot judge any stimulus on its own (see an example below). This means that from the experimenter's point of view, the individual scores for each stimulus are not available; one cannot compute an SV by the usual method of averaging. The difficulty the subject would face can be illustrated by remembering the standard Thurstonian experiment. Let us suppose for instance that the subject is asked to make a psychophysical judgement such as the weight or the warmth of a stimulus. Unless the experimenter assumes that the subject has an implicit (unconscious) but definite standard of judgement for the sense of weight or warmth, he is asking an impossible question to his subject. The same argument holds in the present application. That is, we cannot ask the subject about the sense of similarity concerning a single sentence. As in Thurstone's experiment, we need to supply at least one standard sentence with which the other sentences can be compared (notice the importance
of 'sure' items in the analogical process mentioned in chap. 4). Since it is not tenable to presume what is the optimal standard sentence prior to the experiment, we can regard any sentence as a possible candidate for a standard sentence. Thus, we must try to obtain a scale of values for all possible combinations of paired comparison.

We have seen in the above that one cannot elicit a judgement about a single stimulus in isolation. Thurstone advocates the elicitation of pairwise comparison so that each discrete stimulus has a standard to be compared with. But this creates computational complexities of computation. As we have seen above, the elicitation of pairwise comparison means that one cannot use the standard method of averaging to obtain a central tendency for a single stimulus, because according to the elicitation of pairwise comparison, each scale value needs to be estimated relative to a series of other stimuli. In order to obtain a estimate of a central tendency for a single stimulus in relation to other stimuli, Thurstone proposes a second postulate. That is, Thurstone assumes that the "discriminal difference" ($d_k - d_j$ below) which indicates the difference between a pair of discriminial processes conforms to a normal distribution with variance $\sigma_{d_k-d_j}$.

\begin{equation}
S_k - S_j = X_{jk} \left( \sigma_j^2 + \sigma_k^2 - 2r_{jk} \sigma_j \sigma_k \right)^{1/2}
\end{equation}

where $X_{jk}$ stands for the "normal deviate corresponding to the theoretical proportion of time: stimulus $k$ is judged greater than stimulus $j$"; $r_{jk}$ is the correlation between the pairs of the discriminial process $d_j$ and $d_k$; Torgerson, 1958:161.

![Fig. 1](image1)
![Fig. 2](image2)
$X_{jk}$ (see Fig. 2) represents a mean derived from the standard normal distribution $N(0,1)$ for the discriminable difference between $S_j$ and $S_k$. Since these stimuli form a series (i.e., sequence), all mean values fall along the abscissa of $N(0,1)$ ranging from $-\infty$ to $+\infty$. More important, these values are considered to be divisible in proportion to $\sigma_d \cdot d_j$ units; see, Torgerson, 1958:160-1 and Thurstone, 1932.

Equation (1) above is not solvable as it is, as Torgerson makes clear. Thurstone offers three kinds of simplifying conditions:

A: "the covariance $r_{jk} \sigma_j \sigma_k$ according to Torgerson's notation, 1958:163) is constant for all pairs of stimuli"

B: the correlation terms $r_{jk}$ (above) are all equal.

C: the standard deviation of the distribution of discriminable differences is constant for all pairs of stimuli. Torgerson also regards the categorical judgement as conforming to equation (2):

$$t_g = s_i = \sqrt{x_{jk} \left( \sigma_j^2 + \sigma_k^2 - 2r_{jk} \sigma_j \sigma_k \right)}^{1/2}$$


Two equations (1) & (2) are constructed similarly. This is because Torgerson regards the boundaries ($t_g$) between adjacent categories as behaving like stimuli in Thurstone's model. Thus, the discriminable difference between a given category boundary and a stimulus is assumed by Torgerson to have a normal distribution function.

Our experimental situation belongs to the case Torgerson's equation (2) deals with. According to Torgerson (1958, chapters 9 & 10), equation (2) is not solvable in its complete form. Thus, as in Thurstone, Torgerson considers four simplifying conditions:

Condition A: "the covariance term of equation (2) is constant over all values of $j$ and $g" (1958:207). This amounts to saying that "for all practical purposes, $r_{jg}$ is equal to zero" (idem).

Condition B: "$\sigma_g$ is constant for all values of $g$"; an "the correlation term vanishes". i.e., "the value under the radical in equation (2) is constant for a fixed valu
of j"; 1958:208.

**Condition C:** "\( \sigma_j \) is constant for all stimuli"; and "a correlation terms vanishes" (idem).

**Condition D:** \( \sigma_y, \sigma_j \) and \( r_{yj} \) are all constant.

(Covariance is normally explained:
\[ \text{Cov}(X,Y) = E((X-\mu_1)(Y-\mu_2)) \] where \( \mu = E(X), \sigma^2 = E(X-\mu)^2 \); and correlation coefficient is represented as \( -\frac{\text{Cov}(x_1,x_2)}{\sigma_1 \sigma_2} \).

Since equations (1) & (2) equally deal with two items at a time and each item is assumed to have a normal distribution, the Thurstone-Torgerson method may be more naturally regarded as (dealing with the cases to which we normally apply) a bivariate normal distribution function. It is noteworthy that under this interpretation of the bivariate analysis, Thurstone's main rationale for regarding \( X_{jk} \) as divisible by \( \sigma_d - d_j \) units is tenable. Since the present study follows this solution, it requires some comment.

Equation (3) represents a bivariate normal distribution function and equation (4) indicates that \( X_{jk} \) in equation (1) and \( X_{jg} \) in equation (2) corresponds to \( z \) considered within the framework of the standard bivariate normal distribution function.

\[ f(x, y) = \frac{1}{2 \pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2(1-\rho^2)} \left( \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right) - \frac{2 \rho (x-\mu_1)(y-\mu_2)}{\sigma_1 \sigma_2} \right] \]

\( \rho = 0; \sigma_1 = 1 \)

\[ z = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[ -\frac{1}{2 \sigma^2} (y-x-y^2)^2 \right] \]
It is to be stressed that in eq. 3 all mean values form a linear equation (see Fig. 3). Likewise, when \( f(x, y) \) is considered in relation to the standard normal distribution function, as in eq. 4, the mean values also form a linear equation (see Fig. 4). Even when we interpret the Thurstone-Torgerson method as dealing with the bivariate normal distribution, these two features justify Thurstone's use of linear equation for the purpose of estimating scale values which are simply mean values in our usual terminology. Thurstone's original solution based upon a linear equation yields the following equation, in applying to the categorical judgement as in the present study.

\[
(5) \quad X_{jg} = a_k/a_j (X_{kg}) + (S_k - S_j)/a_j; \quad \text{Torgerson, (1958:221)}.
\]

Eq. 5 is a linear equation and therefore can be evaluated. This equation can be compared with the equation in Fig. 3. In eq. 5 the slope of a line is expressed as the ratio of two dispersion parameters \( (a_{j+1}/a_j) \). This is the same as the linear equation represented in Fig. 3 (i.e., \( \beta \sigma_2/ \sigma_1 \)). The intercept of the line in eq. 5 is equal to the scale value of \( S_k \), since in the Thurstone-Torgerson method the initial value of \( S_j \) and \( a_j \) is determined arbitrarily.
as 0 and 1 respectively. This result is also obtained from the equation presented in Fig. 3. When \( \mu_1 \) is set as 0 and \( \sigma_1 \) as 1, \( y \) is equal to \( \mu_2 \) which is the scale value of \( S_k \). According to Thurstone's use of the linear equation, the other scale values are obtained successively, preserving the linear property. As we have already indicated, \( X_{jk} \) is considered within the framework of the standard normal distribution function. This means that the linear equation of mean values in Fig. 3 we have just considered can be translated into the equation in Fig. 4. In the present application it also means that the scale values for each pair of sentences which the subjects rate varingly along the six categories can be considered as the values falling along a regression line of \( y \) on \( x \) as in eq. 4. This use of the regression line of \( y \) on \( x \) is suggested by Tanaka, 1973. The present SIM adopts this idea.

In the above we have seen that the interpretation of the Thurstone-Torgerson method as the bivariate normal distribution is in keeping with Thurstone's use of a linear equation. We will consider next how this interpretation is not contradictory to Thurstone's or Torgerson's major rationales of the model. In the simplifying condition A mentioned above, the covariance term is assumed to be constant. This means that the correlation coefficient \( r_{jk} \) is zero (see Torgerson, 1958:163). Assuming that \( r_{jk} \) is zero, \( \rho \) is zero. When \( \rho \) is zero in equation 3, we get

\[
(6) \quad f(x,y) = \frac{1}{\sqrt{2\pi \sigma_1 \sigma_2}} \exp \left( -\frac{(x-\mu_1)^2}{2\sigma_1^2} \right) \frac{1}{\sqrt{2\pi \sigma_1 \sigma_2}} \exp \left( -\frac{(y-\mu_2)^2}{2\sigma_2^2} \right) f(x)f(y)
\]

where \( f(x) \) is \( N(\sigma_1, \mu_1) \) and \( f(y) \) is \( N(\sigma_2, \mu_2) \).

That is, when \( \rho = 0 \), the bivariate normal density function \( f(x,y) \) becomes the product of the two normal density functions \( f(x) \) and \( f(y) \). This supports Thurstone's position in that although Torgerson does not represent
the paired discriminial difference as derived from the product of the two normal distribution functions, it is nonetheless reasonable to assume that the respective discriminial processes take normal distribution functions, as Thurstone proposes.

Furthermore, another aspect of the above assumption (\( \rho = 0 \)) is that it is operationally equivalent to the mathematical operation of differentiation; i.e., suppressing the other variable to a minimum, we get a zero correlation. When we differentiate the bivariate function \( f(x,y) \) with respect to either \( x \) or \( y \), we get what is usually called "marginal density functions":

\[
\begin{align*}
(7) \quad (i) & \int_{-\infty}^{\infty} f(x,y) \, dy = f_1(x) \\
(ii) & \int_{-\infty}^{\infty} f(x,y) \, dx = f_2(y)
\end{align*}
\]

These \( f_1(x) \) and \( f_2(y) \) are known to form normal density functions (see, Kogo, 1978:21 & 94). Thus, for this reason also, Thurstone's position that each discriminial process follows a normal distribution is tenable, even when we regard Thurstone's model as essentially bivariate.

The above equations (i) & (ii) in 7 suggest that if we could obtain observed values experimentally with regard to \( \mu_1 \) & \( \sigma_1 \) for \( f_1(x) \), and \( \mu_2 \) & \( \sigma_2 \) for \( f_2(y) \), one can compute the values for covariance and the correlation coefficient in eq. 3, without involving any simplyfying assumptions. Once we obtain these \( \mu_1, \mu_2, \sigma_1 \) & \( \sigma_2 \) separately, one can use the standard formula for covariance \( (x,y) \) and correlation coefficient \( \rho \):

\[
\begin{align*}
(8) \quad (i) \quad & \text{Cov}(x,y) = E \left[ (x - \mu_1)(y - \mu_2) \right] \\
& = E(xy) - \mu_1 \mu_2 \\
& \text{where } E(xy) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y) \, dx \, dy.
\end{align*}
\]

\[
(ii) \quad \rho = \frac{\text{Cov}(x,y)}{\sigma_1 \sigma_2}
\]

Thus, in theory, if the subject's response to the respective stimuli can be independently obtained, it is possible to undertake this bivariate analysis without depending upon
any of the simplifying assumptions offered by Thurstone and Torgerson.

However, as briefly indicated previously in this chapter, it is not possible to elicit from the subject any independent estimates of individual stimuli. Since the judgement of similarity always requires a pair of stimuli; a sense of similarity can not arise from any single item. The same argument holds in the Thurstonian experiments; eg., the subject can judge that this tone is 'louder' than the other, but it is not realistic for us to expect that any subject has an absolute standard of loudness. Since it is not plausible for us to obtain a separate estimate for any individual stimulus, the usual data treatment of bivariate analysis suggested by eq. 6, 7 & 8 does not appear to be possible.

We have argued so far that, since the Thurstone-Torgerson method necessarily involves two stimuli at a time and each discriminant process for these stimuli is assumed to follow a normal distribution function, the present assumption of a bivariate normal distribution function is more natural. Furthermore, we have seen that the postulate of bivariate distribution is not contradictory to the major assumptions of the Thurstone-Torgerson method. That is, by assuming $\rho=0$, as in the Thurstone-Torgerson method, $f(x,y)$ shows that both $f(x)$ and $f(y)$ conform to normal distribution functions. This means that the major assumption concerning discriminant processes holds in view of bivariate normal distribution function. Examining Thurstone's original use of linear equation, we have ascertained a close affinity between the linear equation derived from the bivariate normal distribution function and the linear equation (eq.5) presented by Torgerson as an application of Thurstone's original idea. It must be noted however that the former linear equation represents mean values which fall along a straight line. Thurstone's insight into the use of linear equation can be best understood when we see his model in the light of bivariate normal distribution function. Since it is a linear equation
derived from the bivariate normal distribution function that shows the linearly arranged mean values. What we wish to estimate are those mean values termed "scale values" according to Thurstone's terminology. Moreover, there is another merit in the present reinterpretation of the Thurstone-Torgerson model. That is, the linear equation derived from the bivariate normal distribution function in fact denotes conditional probability:

\[
(9) \quad f(y|x) = \frac{1}{2\pi\sigma_2\sqrt{1-\rho^2}} \exp \left[ \frac{-1}{2(1-\rho^2)} \left( \frac{y-\mu_2 - \rho \frac{\sigma_2}{\sigma_1} (x_1-\mu_1)}{\sigma_2} \right)^2 \right]
\]

(This result follows from eq. 3, 6 & 7(i).)

The conditional density function \(f(y|x)\) forms again a normal distribution with mean \((\mu_2 + \rho \sigma_2/\sigma_1 (x - \mu_1))\) and variance \((\sigma_2^2 (1-\rho^2))\) -- see Fig. 3. This series of mean values fall along a straight line, as we have seen previously. When we assume a space for the standard normal distribution, as in Fig. 4, we get the present solution. It is to be stressed here that once we determine which pair of sentences can be best regarded as an arbitrary origin (a zero point) in the similarity continuum, the other scale values can be estimated successively along that continuum. Because the linear equation of mean values is derived from the conditional density function coming from the bivariate normal distribution function, Thurstonian successive scaling based upon the previous values appears to be fully justified.

The actual procedure of calculation is summarized below.

1. Construct a frequency matrix (F) consisting of rated scores.
2. Convert the F matrix into the matrix of cumulative probability (P).
3. Rearrange the rows of P matrix; by summing matrix P over \(T_g\) (categories) for each row, we arrange the rows in the order of increasing magnitudes of the sums (see Torgerson, 1958:221-7.)
4. Convert the P matrix into the matrix of normal
deviates \( X \).

(5) Compute scale values for each pair using Thurstone's equation. We also use "normal equations" to estimate the regression line of \( y \) on \( x \).

(6) Compute \( t \) (category boundaries) on the assumption of Torgerson's simplifying condition \( B \). But \( g_i \) is estimated afterwards.

We have seen in the above that Thurstone's use of linear equation can be interpreted in relation to a bivariate normal distribution function. In addition to the merits of the present re-interpretation mentioned above, another merit can be indicated. That relates to our discussion in §7.0. Each stimulus gives rise to a discriminial process, which each stimulus possesses inherently and to which the subject is attending. These criterial attributes may or may not be singular and they may be poorly or explicitly structured. Since the experimenter cannot judge which class of these possible situations a given group of subjects belongs to, it would be safe to rely on the method of measurement which is as inclusive as possible. For this purpose, the present method is effective, since function \( f \) in \( f(x,y) \) does not make any explicit assumption as to the substantial nature of the function; it only specifies the shape of the response distribution. The substance of function can be any predicate for which there is a mathematical representing function (eg., predicate, 'similar'); for the relationship between predicates and functions, see Kleene, 1952. This may account for the reason why the Thurstone-Torgerson method has been widely applied to the measurement of attitudes (see, chap. 5), since they involve highly intricate predicates which may be represented as various composite functions. It thus suggests that the Thurstone-Torgerson method is more general than expected. By looking at the model in light of the bivariate distribution, one can understand at least a partial reason why the model is eligible in many situations. This is particularly relevant in this research. The outside experimenter cannot predetermine what is the basis of the subject's similarity.
judgement. It could be a singulary predicate or some complex predicate leading to some composite function as its representing function (the term from Gödel, 1930). This may include the predicate 'similar'. By postulating this 'similarity' predicate as function \( f(x,y) \) for its representing function, it enables us to be prepared for either of two possible situations: (1) the judgement of NL similarity is derived from an explicit type of composite function consisting of some salient semantic features; (2) the judgement is derived from the undifferenciated intuitive 'feel' \( f(x) \) which is still represented as predicate 'similar'.

§ 7.2 Experiments

Two experiments are conducted to find out whether the derived scale values (SV) and the estimation of category boundaries \( t_g \) are reasonable. Since these experiments involve separate hypotheses, the hypotheses are presented in separate sections, §7.2.1 & §7.2.4. The results concerning these hypotheses are presented in §7.2.3 & §7.2.5.

§ 7.2.1 hypothesis 1

We have seen above that, although Thurstone's model involves bivariate functions, Thurstone regards the distribution of the difference between successive pairs as normally distributed, so that several simplifying computational procedures are made possible. Furthermore, since we are dealing with consecutive pairs, the basic rationale of the model is to the consideration of the standard normal distribution corresponding to such successive pairs. Hypothesis 1 examines whether this assumption is reasonable. Specifically, we examine whether the scale values derived from the assumption of standard normal distribution offer reasonable measures in relation to \( t_g \). This can be achieved by comparing the prediction derived from the model with the experimental observation described below. The extent to which the prediction corresponds to the
experimentally observed values determines the degree in which the model adopted for our present purpose is adequate.

In this experiment, we present the subjects with eight pairs of sentences and ask them to respond whether the pair is within the range of criteria given. The criteria have the lower and upper judgemental limits. In this sense, the present experiment also tests whether the judgemental limits calculated by SIM are adequate or not. The extent to which our predictions correspond to experimental observations determines the adequacy of our rationale.

As we have seen above, we assume that the variances of category boundaries are constant. The distribution of the difference between the successive pairs, we assume, takes a standard normal distribution. Under these assumptions, when the subjects are asked to respond whether a pair of NL sentences falls within the range of prespecified criterion (eg., 'identical', 'very different', etc.), the subjects will make either negative or affirmative responses, and the proportion of negative or positive responses can be predicted by the equations.

The individual response depends upon the subjective scale value(x) each individual conceives. The scale value(SV) derived from the Thurstone-Torgerson method is simply an average across individually differing scale values (x); and it is thus reasonable to assume that individual subjects have their own x. These individually differing x's serve as decision points for the subjects in making their responses. If this subjective SV is within the range of the criterion, it will lead to a positive response. Whereas when x exceeds the criterial range, it will lead to a negative response. Since the Thurstone-Torgerson method specifies the SV and discriminant dispersion(DD) for each pair of sentences, one can determine the probability density function \( \phi(x) \) for the pair. This enables us to predict the proportion of the subjects who would respond positively or negatively. The proportion of positive responses is the integral of \( \phi(x) \) from \( t_1 \) to
(9) \[ \text{Pr(pos)} = \int_{t_1}^{t_j} \phi(x) \, dx = \Phi(t_1) - \Phi(t_j). \] (see Fig.6)

Similarly, the proportion of negative responses is the integral of \( \phi(x) \) from \( t_1 \) to \( \infty \) plus the integral of \( \phi(x) \) from \( t_j \) to \( -\infty \):

\[ \int_{t_1}^{\infty} \phi(x) \, dx + \int_{t_j}^{\infty} \phi(x) \, dx = 1 - \Phi(t_j) + \Phi(t_1). \] (see Fig.7)

Eight pairs of sentences with kowasu were chosen for this experiment. These pairs were divided into two: the first four pairs were presented with the criterion "very different" and the remainders with the criterion "closely related". The former criterion falls between two category boundaries, \( t_4 \) & \( t_5 \). The latter criterion falls between \( t_1 \) & \( t_2 \). Thus, we can get the following predicted proportion for the eight paired sentences:

\[
\text{closely related: } T_1 - T_2
\]

\[
\begin{array}{cccc}
\text{T}_1 & \text{T}_1 & \text{T}_1 & \text{T}_1 \\
0.5916 & 1.4633 & 1.9488 & 2.4710 & 3.0925 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{closely related} & \text{SV} & \text{DD} & \text{predicted} \\
\text{mokei/taicho} & 2.6908 & 0.4163 & \text{yes} & 0.06 & 39.94 \\
\text{mokei/kikaku} & 2.6618 & 0.5411 & \text{yes} & 0.54 & 39.46 \\
\text{mokei/taisei} & 2.3252 & 0.6177 & \text{yes} & 3.13 & 36.87 \\
\text{okimono/kibo} & 3.1337 & 0.6590 & \text{yes} & 0.00 & 40.00 \\
\end{array}
\]

Table 2

238
very different: \( T_4 - T_5 \)

![Graph showing very different](image_url)

<table>
<thead>
<tr>
<th>very different</th>
<th>SV</th>
<th>DD</th>
<th>predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>taicho/okimono</td>
<td>2.7321</td>
<td>0.4589</td>
<td>yes</td>
</tr>
<tr>
<td>taicho/tofu</td>
<td>2.6895</td>
<td>0.5934</td>
<td>39.24</td>
</tr>
<tr>
<td>mokei/kibo</td>
<td>3.0135</td>
<td>0.5432</td>
<td>39.91</td>
</tr>
<tr>
<td>okimono/taisei</td>
<td>2.5106</td>
<td>0.6900</td>
<td>37.54</td>
</tr>
</tbody>
</table>

Table 3

§ 7.2.2 administration: experiment 1

44 first-year students at Ehime University took part in the experiment. Of the 11 sentences with kowasu which were chosen in chap. 5, 7 sentences were selected for initially this experiment. As explained below in §7.2.4, due to the restrictions of the tachistoscope presentation, only four paired sentences (see Table 2) were presented to the subjects with the criterion "closely related" and the other four pairs, with the criterion "very different". The paired sentences with these criteria were written in the centre of B4 sheets. Each sheet was presented for one second on a standard tachistoscope. The subjects pressed the key when they made their decision. The order of presentation was counterbalanced. The subjects had two practice sessions.

After the experiment the subjects were provided with questionnaires for "Repression and sensitization scale".
The questionnaires were from a standardized version composed by Togari (1964). This special precaution was taken on the advice of the Research Laboratory at Ehime University. In accordance with their results (Togari, 1977), two students were diagnosed as repressors (2 points), whereas the other two students appeared to be sensitizers (43 points and 44 points respectively). Since they were either very hesitant to respond or too prompt but inconsistent, these 4 students were omitted from the analysis of the following two experimental results.

§ 7.2.3 result: experiment 1

Tables 4 & 5 represent the results of experiment 1. Table 4 refers to the 4 paired sentences with criterion "closely related", and table 5, those with criterion "very different." Irrespective of the difference of the criterial categories, the observed proportion of subjects corresponds to the predicted proportion.

closely related

<table>
<thead>
<tr>
<th></th>
<th>predicted</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>mokei/taicho</td>
<td>0.06</td>
<td>39.94</td>
</tr>
<tr>
<td>mokei/kikaku</td>
<td>0.54</td>
<td>39.46</td>
</tr>
<tr>
<td>mokei/taisei</td>
<td>3.13</td>
<td>36.87</td>
</tr>
<tr>
<td>okimono/kibo</td>
<td>0.00</td>
<td>40.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>predicted</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>mokkei/taicho</td>
<td>39.89</td>
<td>0.11</td>
</tr>
<tr>
<td>mokkei/tofu</td>
<td>39.24</td>
<td>0.76</td>
</tr>
<tr>
<td>mokkei/kibo</td>
<td>39.91</td>
<td>0.09</td>
</tr>
<tr>
<td>okimono/taisei</td>
<td>37.54</td>
<td>2.46</td>
</tr>
</tbody>
</table>

Table 4

very different

<table>
<thead>
<tr>
<th></th>
<th>predicted</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>mokkei/taicho</td>
<td>39.89</td>
<td>0.11</td>
</tr>
<tr>
<td>mokkei/tofu</td>
<td>39.24</td>
<td>0.76</td>
</tr>
<tr>
<td>mokkei/kibo</td>
<td>39.91</td>
<td>0.09</td>
</tr>
<tr>
<td>okimono/taisei</td>
<td>37.54</td>
<td>2.46</td>
</tr>
</tbody>
</table>

Table 5
According to these results, both SV and Tg obtained by the Thurstone-Torgerson method appear to be reasonable in reflecting the participants' subjective semantic distance between the paired sentences.

§ 7.2.4 hypothesis 2

Given the criterion of similarity judgement and a pair of sentences, the psychologically plausible estimates of SV and DD can predict fairly accurately the time required for such a decision process. This reaction time experiment is also a standard technique by which one can demonstrate experimental evidence for a model which yields such numerical information as SV and DD.

In the present situation, this reaction time experiment involves at least three temporal stages: (1) reading time during which the subjects read a pair of sentences and a given criterion; (2) comparison time during which the subjects judge a pair of sentences according to the prespecified criterion; (3) response time in which the subjects press the key, indicating their negative or positive responses. This experiment is mainly concerned with the second stage of comparison time.

\[
T = T_1 + T_2 + T_3
\]

\(T\) : total reaction time

\(T_1\) : (1) reading time

\(T_2\) : (2) comparison time

\(T_3\) : (3) response time

Comparison time \((T_2)\) may be expressed as the function of the distance between an individual's SV and the location of a criterial category boundary (see Fig. 8). This at least recaptures our intuition. If an individual's SV is very distant from the upper or lower limit of a criterion, he can make a quick decision; on the other hand, if his SV is close to the upper or lower limit, he will be slow to respond. In order to express this straightforward intuition mathematically, one only needs to follow a standard practice in mathematical psychology;
the latency mechanism operating over continuous time rather than over discrete trials has been regarded as exponential functions or more generally gamma distribution; Luce, 1963, Thomas, 1971, Atkinson & Juola, 1973, 1974, etc. The use of exponential functions appear to be tenable, since the mechanism of neurons in the nervous system, particularly the transmission of information, is often represented by these functions; Lubinsky & Pozin, 1965, Aida, 1967, 1959, etc. Following Atkinson & Juola (1974), this study regards comparison time as the exponential function of the distance between an individual's SV and the location of a criterial category boundary:

\[(11) \quad T_2 = e^{-\beta(x-T_c)}\]

The above equation can be amended to take into account the effects of practice over trials. But as in Atkinson & Juola (1973), this is avoided by restricting the analysis to the later trials in which performance normally reaches some asymptotic level. In the present experiment, we assume that the above equation is sufficient to represent...
our straightforward intuition concerning the relationship between internalized scale values and judgmental processes. That is, the subjects consult their subjective lexicon about the sense of semantic distance between the pairs; and these subjective assessments are directly reflected in their decision making: the remoter the subjective SV is to a given category boundary, the more readily they can respond, whereas the closer the SV is to the boundary, the more hesitant the response, leading to the longer latency.

What we wish to demonstrate is that the estimates of SV obtained by the Thurstone-Torgerson method are reasonable. This requires two things. First, eq. 11 is assumed to be reasonable. Second, the assumption that varying SVs directly reflect the process of comparison needs to be tenable. The second assumption is the main issue here, and we assume the experimental situation does not involve any other biasing variables. This assumption is accepted here without performing any further experiment. The first assumption suggests that we must be able to obtain the optimal estimates for parameters 'd' & 'B' in the equation. In order to achieve this, there are again two methods (A) & (B) below.

\[(A) \text{ As in Atkinson & Juola (1973 & 1974), parameter estimates are selected so that the sum of the squared deviations between the 12 data points (see Table 9) and the theoretical predictions can be minimized.}\]

\[(12) \text{ RMSD} = \left( \frac{1}{N} \sum_{i=1}^{N} n_i (t_{p,i} - t_{0,i})^2 \right)^{1/2}\]

"where N = the total number of observations; i = an index over the 12 data points shown Table 9; n_i = the number of observations determining data point i; t_{p,i} = predicted response latency for data point i; and t_{0,i} = observed response latency for data point i."


As in Atkinson & Juola (1974), we can assume further that
apart from the factor of the semantic distance between an individual's SV and a given criterial category boundary, the 4 pairs of sentences dealt with here should not involve any other variables. If we assume this position, the mean square root deviation can be used to determine \( \alpha \) and \( \beta \) on the basis of the 12 estimates of parameters.

(B) Alternatively, since we assume that the SVs are arranged linearly according to Thurstone's original idea, one can use normal equations rather than the root mean square deviation: see § 7.1 for Thurstone's linear assumption. In eq.12, parameter estimates are selected to minimize the sum of the squared deviations. For the purpose of minimization, method (B) adopts the classical normal equation which is used in the calculation of regression line (Fig. 4). This idea appears to be related to Thurstone's linear assumption (see Fig. 3 & 4). As Thurstone assumes SV to be linearly allocated, so is \( (SV - T_g) \).

As we have seen above, the total time required for the completion of the experimental task breaks down into three consecutive stages: \( T_1 \) (reading time); \( T_2 \) (comparison time); \( T_3 \) (response time). If the time needed to read and to respond by pressing the key differs according to the four paired sentences, this may distort the results. Since any irrelevant factors ought to be kept as constant as possible, it is necessary for us to check that the sum of \( T_1 \) and \( T_3 \) does not show any significant difference across these pairs. The following consideration was given to this. The ease of reading each word is determined by a number of factors such as the number of strokes in Chinese characters, the size of a Chinese character, the spacing between characters, the intensity of illumination, the distance between characters and the position of the subject, etc. Apart from the first factor, the others are kept constant, when using a tachistoscope. In the present experiment, the size of a character is \( 4 \times 4 \) cm\(^2\); the width of a line of the stroke of a character is 2 mm; the intensity of illumination is about 230 lux; the value
is about 9.1; the spacing between lines is 7mm; the spacing between letters is 5 mm. Since the number of strokes in Chinese characters is different, the total time to complete reading and pressing of a key was independently measured and examined to see whether there was any significant difference among the four pairs presented.

In summary, in this experiment the time required for $T_1$ (reading time) and $T_3$ (response time), in which the subject merely read the stimuli and pressed the key, is separately measured. We then subtract $T_1$ and $T_3$ from the total time $T_1 + T_2 + T_3$ so that we only deal with only $T_2$. On the basis of this $T_2$, we will estimate parameters 'a' & 'b' in eq. 11 in two ways: one following Atkinson & Juola (1974) and the other taking into account Thurstone's linear assumption concerning SV. We then compute the predicted comparison time on the basis of eq. 11. This equation is regarded as representing our simple intuition concerning the relationship between SV and the decision process. The closer the SV is to a given category boundary, the longer it takes for the subject to arrive at his decision. On the other hand, the remoter the SV is to a given category boundary, the quicker the decision is. That is, we assume that $T_2$ can be represented as an exponential function of the semantic distance between SV and a given category boundary.

This reaction time experiment is also intended to see whether the judgemental limits calculated on the basis of SIM are adequate. By examining the accuracy of $T_g$, we achieve this experimental aim.

§ 7.2.5 administration: experiment 2

The same students as in the previous experiment participated in this experiment. The stimuli were written on B4 sheets. The size of characters, the width of each stroke of print and the spacing between letters are reported in the previous section. The same procedure as in experiment 1 was employed except for the following. Of the eight paired
sentences, four pairs with criterion "closely related" were regarded as stimuli and the other four pairs were used as distractors. The order of presentation was counterbalanced, as in experiment 1. Each pair was presented for 0.8 of a second. The process was repeated five times. The reaction times were recorded on a paper punch system.

§ 7.2.6 results and discussion: experiment 2

As Table 7 in Appendices indicates, trials 3 to 5 reached a satisfactory asymptotic level. It suggests that we do not have to take any practice effects into consideration, as in Atkinson & Juola's study.

<table>
<thead>
<tr>
<th></th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>mokei/taicho</td>
<td>M 823.10</td>
<td>828.38</td>
<td>829.50</td>
</tr>
<tr>
<td></td>
<td>SD 125.71</td>
<td>130.47</td>
<td>130.15</td>
</tr>
<tr>
<td>mokei/taisei</td>
<td>M 906.31</td>
<td>903.72</td>
<td>901.23</td>
</tr>
<tr>
<td></td>
<td>SD 100.15</td>
<td>101.68</td>
<td>103.68</td>
</tr>
<tr>
<td>mokei/kikaku</td>
<td>M 863.29</td>
<td>860.66</td>
<td>855.87</td>
</tr>
<tr>
<td></td>
<td>SD 144.03</td>
<td>142.01</td>
<td>141.71</td>
</tr>
<tr>
<td>okimono/kibo</td>
<td>M 769.25</td>
<td>772.90</td>
<td>769.33</td>
</tr>
<tr>
<td></td>
<td>SD 139.27</td>
<td>143.07</td>
<td>135.73</td>
</tr>
</tbody>
</table>

Table 7

Further, Table 8 indicates that there is no statistically significant difference among the times required to read and to respond to each of the paired sentences, i.e., \( T_1 \) (Reading Time) + \( T_3 \) (Response Time) shows no significant difference. Thus, by subtracting \( T_1 \) and \( T_3 \) from the total time ( \( T_1 + T_2 + T_3 \) ), we get \( T_2 \) (comparison time) during which the subjects have arrived at a similarity judgement for each pair.

\[ T_1: \text{reading time} \]
\[ T_2: \text{comparison time (discriminal process)} \]
\[ T_3: \text{response time} \]
No of Strokes | M    | SD    | No of Subjects
-------------|------|-------|----------------
      | T1 + T3 |
mokei/taicho | 45   | 394.75 | 96.11 | 40 |
mokei/taisei | 38   | 412.68 | 75.89 | 37 |
mokei/kikaku | 38   | 416.87 | 106.16 | 39 |
okimono/kibo | 40   | 404.25 | 98.84 | 40 |
( F=0.4145 df=3 N.S. )

Table 8

On the basis of T2, we estimate parameters α & β in eq. 11. Table 9 presents the result of parameter estimations per trial for each pair.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Inα</th>
<th>Inβ</th>
</tr>
</thead>
</table>
mokei/taicho | 3  | 6.5511 | 0.4001 |
    | 4  | 6.5511 | 0.3901 |
    | 5  | 6.5511 | 0.3880 |
mokei/taisei | 3  | 6.5511 | 0.4053 |
    | 4  | 6.5511 | 0.4114 |
    | 5  | 6.5511 | 0.4173 |
mokei/kikaku | 3  | 6.5813 | 0.4006 |
    | 4  | 6.5511 | 0.3803 |
    | 5  | 6.5511 | 0.3891 |
okimono/kibo | 3  | 6.5511 | 0.4374 |
    | 4  | 6.5511 | 0.4304 |
    | 5  | 6.5511 | 0.4373 |

Table 9

Based upon the estimates of parameters α & β, we can compute the predictable reaction times in the two methods mentioned above. Method (A) is Atkinson & Juola’s "mean squared deviation" (see eq. 12). Method (B) incorporates Thurs- tone’s linear assumption (in the calculation). That is, parameter β is linearly related to (SV - T9). Table 10 compares predicted T2 with observed T2.

<table>
<thead>
<tr>
<th>prediction (1)</th>
<th>prediction (2)</th>
<th>observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>426</td>
<td>433</td>
<td>432</td>
</tr>
<tr>
<td>497</td>
<td>503</td>
<td>491</td>
</tr>
<tr>
<td>431</td>
<td>442</td>
<td>443</td>
</tr>
<tr>
<td>382</td>
<td>364</td>
<td>366</td>
</tr>
</tbody>
</table>

Table 10
According to Table 10, method (B) yields a better fit to the observed latency than method (A) does. This suggests that Thurstone's linear assumption appears to be more relevant in the present experiment than the method of "mean squared deviation".

§ 7.3 conclusion

The two experiments were conducted to see whether the present application of the Thurstone-Torgerson method would be suitable for the experimental materials used. §7.2.3 & §7.2.6 present the results of the two experiments. The discussion shows that the present application of this SIM is tenable in terms of (1) the ratio of positive/negative responses and (2) the latency with respect to Comparison Time (T_2) in Diagram 1. Assuming that the SVs derived from SIM were given, the above numerical predictions were made. We have seen that the predictions were experimentally born out, indicating that the present application of the Thurstone-Torgerson method appears to be appropriate for the purpose of estimating the judgemental limit and the 'feel' of similarity between the paired items.

The above results are relative to the experimental items used. The two experiments dealt with only four pairs of sentences with the four T_9s. This limitation is due to the following two reasons. (1) The other paired sentences are likely to cause Reading Time (T_1) in Diagram 1 to differ significantly. This is because the number of strokes in Chinese characters differs a great deal (from 37 to 45 strokes) -- distorting factors affecting Reading Time, see §7.2.4. As we have seen above, if this Reading Time differed significantly in each pair, we would not have made an accurate prediction of the latency for the discrimininal process (T_2). For this reason, only the 4 pairs were chosen. (2) In experiment 2, some pairs were avoided for a different reason. Some pairs were expected to take a Laplace distribution function (see Wilde, 1964). For instance, a pair (taisei & kikaku) with the criterion 'some relationship' is likely to follow a Laplace function.
According to Wilde (1964), the estimation of parameters in Laplace functions is difficult to make, as Fig. 9 illustrates.

With these limitations of the present experiments in mind, one might tentatively say that the application of the Thurstone-Torgerson method in the present research is justifiable.

§ 7.4 the application of SIM to the analysis of judgement data

In this section, we discuss the relevance of SIM for the analysis of judgement data (chap. 10), in view of the mathematical characterization shown above. The relevance of SIM for that purpose was experimentally examined in chap. 4. And in this chapter we approached SIM differently and tested the accuracy of SVs and Ts derived from SIM. Further, while in chap. 4 the mathematical nature was not fully detailed and this chapter supplied information necessary for the further discussion below. The above slightly technical characterization of SIM enables us to appreciate the greater relevance of SIM for the analysis of translatability judgement data.

In the experiments (chap. 7), we have assumed that the subject conceives his subjective SVs, when presented with a specific pair. We instructed the subject to judge whether the criterion given (e.g., 'very different', etc.) was satisfied by the specific pair; i.e., he was asked to respond negatively or positively to the criterion. This experimental situation is analogous to the elicitation of translatability judgement in chap. 4 & 10 for the following sense. On eliciting the judgement data from subjects, we assume that they conceive their own SVs for a given pair of sentences. This is because each subject is assumed to have a particular 'sure' item in mind (i.e.,
he is confident of the translatability of this item. And his judgement will be based upon the paired comparison between his 'sure' item and the item whose translatability he is judging (for the definition of 'sure' item, see chap. 4). In chap. 10, the youngest group participated in the experiment is given an experimental 'sure' item. And the older subjects have learned some of the items on the list at school long before the experimental sessions. In this sense, all subjects should have their 'sure' item(s). Because the paired comparison is involved in the judgemental process, the 'feel' of semantic distances between the 'sure' item and other items play an important role in the judgemental process. For this reason, we require an adequate measurement of this 'feel' of semantic distance.

We can assume further that the subject may conceive some judgemental limit in the process (see chap. 3). This judgemental limit may be explained in comparison with the experimental criterion given to the subjects (see above). In the above experiment, the subject responded "yes", if his SV was within the range of the criterion given; and if it exceeded the range, he would respond negatively. Likewise, in making a translatability judgement, the subject may be assumed to conceive his judgemental limit analogous to the experimental criterion above. If his SV is within the judgemental limit, he will regard the item as translatable. If his SV exceeds the judgemental limit, he will reject the item as untranslatable. Thus, if the idea of judgemental limit holds true in the actual experiments in chap. 10, we can recognize a similar function in the experimentally supplied criteria above and this 'hypothetical' judgemental limit of the translatability judgement.

The above hypothetical argument will be experimentally tested in chap. 10. I will take a liberty of elaborating the judgemental limit here in comparison to the experimental criteria above, since the latter bears some analogy with the former. The subject may conceive this judgemental
limit conceptually or notionally. The experimental criteria were supplied to the subjects in notional terms; e.g., 'very different' or 'closely related'. The judgemental limit may be similarly conceived by the subject (in chap.10) as 'some relationship' or 'closely related'. If so, this conceptual judgemental limit functionally corresponds to the criterial range or the category boundary \( T_g \). The adequacy of the latter was experimentally demonstrated in this chapter. In this sense, it would be justified to calculate the judgemental limit, based on SIM. The actual procedure for this calculation was explained in experiment 1 of this chapter.

The above judgemental limit may be regarded as representing "inductive limit" in a mathematical term (for the definition of "inductive limit", see Takeuti, 1978 & Matsumoto, 1980). The discussion could be technical, but the mathematical details for this purpose are sufficiently supplied in this chapter and Takeuti's definition accords with the present understanding of judgemental limit, when SIM is regarded as the simplified version of bivariate analysis. The close relationship between SIM and bivariate analysis was discussed in this chapter. Here, we will discuss only the advantage of introducing this mathematical terminology into the present view of IL acquisition. This issue relates to the useful distinction between hypothesis formation and hypothesis testing. The inductive limit relates to the former and "projective limit" relates to the latter.

In §1.1, I referred to Corder (1976) and we have seen that there are three stages in language acquisition: (i) data-processing, (ii) hypothesis formation and (iii) hypothesis testing. When the subject conceives the judgemental limit for the first time, it is a case of inductive limit. The learner induces the limit of application of a TL word from a particular example. When the learner uses the same judgemental limit on later occasions, it is a case of projective limit. That is, he deduces a conclusion of the translatability of a NL
item whose translatability he is not sure of, by using the judgemental limit which he has acquired on the previous occasion. The present view is also detailed in §3.3 and §3.4.

The three stages of language acquisition are observable among L₁ children (chap. 3). We have also seen that Rosch's prototype model of lexical structure whose relevance in L₁ acquisition is experimentally proved has one ambiguous aspect. The introduction of judgemental limit appeared to clarify this ambiguity. The present discussion in connection with SIM also suggests that the judgemental limit operates inductively as well as deductively in the process of hypothesis formation and hypothesis testing. It is noteworthy that SIM enables us to evaluate the actual value for a specific judgemental limit, as the experiments in this chapter have shown.

We have seen in this chapter that Thurstone's method may be regarded as the economical way of evaluating the complex bivariate event. I have explained how Thurstone's equations (interpreted by Torgerson) relate to the bivariate normal distribution function. This function used in the manner of SIM relates to Takeuti's definitions of inductive limit and projective limit (Takeuti, 1978). The illustrative account of this relationship is given in appendix 5. Assuming that the account outlined in appendix 5 holds true, the present use of terms 'inductive limit' and 'projective limit' will be mathematically well-grounded at the same time.
Chapter 8 pilot study (5); experimenting the hypothesis of compound vs coordinate bilinguals

§ 8.0 purpose

In this chapter the old controversy over compound vs coordinate bilinguals is taken up. As we see below, this dispute has been left inconclusive, and it is interesting to figure out some partial reasons why it has been inconclusive. In addition to this research interest, the hypothesis is relevant in this research. The controversy can be summarized as the issue of whether interlanguage continua support two separate language systems in cognition, one for \( L_1 \) and the other for \( L_2 \) (i.e., coordinate bilinguals) or a single language system common to the two languages (i.e., compound bilinguals): see Fig. 1 & 2. If the subjects who took part in the present research were pure coordinates, native language interference would simply not occur at all. Nor would it be meaningful to use as experimental materials their native language. If learners possess an extreme form coordinate language system, it is unlikely that such learners can judge the translatability on the basis of their native language.

An examination of this hypothesis may be relevant to the theoretical issues in interlanguage. Particularly, Dulay and Burt's recreation vs restructuring hypotheses or Anderson's denativization and relexicalization may also be related to this old dichotomy of compound and coordinate bilinguals (see chap.3). That is, here is an issue of whether, as a result of recreation, restructuring, denativization or relexicalization, two distinct language systems, one for \( L_1 \) and the other for interlanguage, are formed in cognition or simply a single language system is formed irrespective of input languages. Specifically, the hypothesis of recreation process supports the coordinate language system of interlanguage. Whereas, the restructuring hypothesis prefers the compound language system of interlanguage.

253
Coordinate bilinguals by definition also pose a theoretical problem in the present research. If the notion of a distinct language system for each language is correct, it would be highly difficult for any form of analogical inference based upon NL to occur. It would suggest that Ausubel's meaningful learning is unlikely to occur; Quine's notion of innate similarity standard might also be questionable among coordinate bilinguals (see chapters 2 & 3).

§ 8.1 overview

This old dispute goes back to Osgood and Ervin(1954). Osgood & Ervin's distinction between compounds and coordinates appears to follow the stimulus-response paradigm modified by from representational mediation processes. Learning a language is regarded as a process of habit formation and a gradual strengthening of associations between language behaviour and the context of situation in which a given language behaviour takes place. Thus, if a learner who is acquiring two languages at the same time happens to be immersed in two different acquisitional contexts, a separate development of representational mediation processes for each of these languages is hypothesized by Osgood and Ervin (coordinate system): see Fig. 1. For instance, if a learner has learned to speak one language at home and the other in school or at work, he will be a coordinate by definition and two different representational processes for the two languages will be developed in his cognition. On the other hand, if he learns two languages in the same environment, the two representational mediation processes will be unified as a compound system; see Fig. 2.

In this hypothesis, the difference of acquisitional contexts has been regarded as the main source in creating the two kinds of bilinguals, but the hypothesis later included other factors such as different manners of acquisition (Lambert et. al 1958 & Lambert,1969): different teaching methods (e.g., grammar/translation method vs audio-lingual
method); the age of acquisition i.e., before or after puberty (Geneseese et. al 1978); and attitudinal differences toward the cultures of the target languages (Jakobovits, 1970). More recently the lateralization process of the right-hemisphere in the brain is thought to support the coordinate system (Albert & Obler, 1978). This interpretation is not discussed below, since it is beyond the scope of this research.

It may be noted that coordinate bilinguals (Fig. 1) would find any translation task impossible, since there is no route which connects the representation of language A with that of language B. This is implausible, since any bilingual can perform some translation of one language into the other. Osgood and Ervin therefore include three possible mediation processes to account for this translation task in the case of coordinates:
This modification clears two areas of conflict.

(1) It shows that coordinates can perform some translation.

(2) It also explains the fact that translation requires more time than monolingual speech, since the modification specifies the longer route for the translation task.

The modification is however less convincing in the following two respects.

(A) The modification provides a longer route only for the coordinate system. Common sense tells us that any form of translation is a highly conscious task, requiring some objective assessment that the given original and its translation are really equivalent. Translation involves some elaborate comparison of the two not only at every level of sentential hierarchy (i.e., word level, phrase level, etc.) but also some global comparison in terms of syntax and semantics. Thus, it is likely that both coordinates and compounds would find translation equally complex and time-consuming. In this sense the above modification appears to be unsatisfactory, since it only explains the difficulties coordinates would face in translation.

(B) In the above modification the longer route is established, mainly because the hypothesis assumes a separate language system. The complexity of translation may be accounted for in other ways than a priori specification of the two distinct language systems in cognition. For instance, it is true to say that for some lexes there is no one-to-one correspondence between two languages; one lexical item may be found its equivalence only in phrasal expressions or even sentential paraphrases. The fact that any translation is elaborate and time-consuming may thus be due to this different lexicalization processes \( (S_m \to R_A \text{ or } R_B) \) which is not regarded as the crucial part of the hypothesis by Osgood and Ervin (see Fig. 1 - 3). We must also bear in mind that the number of lexes required to express one concept does not necessarily mean this concept is conceived by a learner to be separately represented in cognition. The lexical difference on the surface level may not have anything to
do with the deeper understanding of semantic content (e.g., a circumlocution 'four-legged domestic animal which normally barks at strangers' would be understood by most language-learners that the paraphrase refers to a dog).

Thus, the above modification is arguable in that

(a) compounds would find translation as complex and time-consuming as coordinates and that

(b) translation may only involve the peripheral part of lexicalization which is represented by the hypothesis as entirely equivalent in the two kinds of bilingual language systems (see the route $S_m$ to $R_B$ or $R_A$ in Fig.s 1 & 2).

For this reason, the above modification is not considered below. We will concentrate on the distinctive feature of the hypothesis which differentiates the language systems, i.e., whether interlanguage involves two separate representations for two languages, as in the coordinate language system, or whether interlanguage is to be viewed as a unified language system as in the compound language system.

§ 8.1.1 Specification of the previous experimental area

Before discussing some experimental evidence concerning the hypothesis, one needs to know what linguistic area is meant to be represented by this single or double system. It appears that although the originators use an inclusive term 'language system', the single or double "representational mediation process" roughly refers to semantics for the following reason. The sound patterns a learner hears or the visual patterns a learner recognizes are represented as signs $S_A$ and $S_B$ equally in the two systems (see Fig. 1 & 2). Likewise, the sound or visual patterns a learner produces are represented as $R_A$ and $R_B$ equally in the two systems. This suggests that the hypothesis is not concerned with the differences in two languages which can be described by orthography and acoustic or articulatory phonetics. What makes a distinction between compound and coordinate systems is whether the middle part
of above figures represented by the dotted lines involves parallel routes or a unified single route. One may wonder whether the idea of double syntax is plausible. First of all, it must be noted that an S-R paradigm usually does not take the generative nature of syntax into account as an internalized system of languages; and an S-R paradigm tends to stress the importance of imitation and rote-learning as the major sources of language learning. Since generative grammars aim to establish language-independent (and ideally universal) syntax, it is not tenable from this point of view to claim that there are two kinds of syntax for two languages. It appears that the notion of coordinate system is counter-intuitive in terms of the standard model of generative grammar. Since Osgood and Ervin's hypothesis was proposed around the time when the initial theory of generative grammar was proposed by Chomsky, we will not examine this aspect according to the further development of generative grammars. Thus, we have some reason to disassociate the hypothesis from syntax. It would be slightly more reasonable to regard the hypothesis as dealing with deep semantic coding relating to comprehending an utterance and preparing a speech. Even when we limit the scope of the hypothesis to this area of semantics, it is still confounding. It is difficult to think that a learner develops two separate denotations for the same entity, e.g., apple and pomme. This suggests that the hypothesis is confined further to the subjective and psychological part of semantics a learner possesses.

Two kinds of experimental techniques were often adopted to obtain experimental evidence that the distinction between compound and coordinate systems is meaningful and real. By looking at these major experimental techniques, one can get a clear idea of what is postulated as a single or double system. These are the "semantic differential method" (e.g., Lambert et al., 1958) and the "semantic satiation method" (Lambert & Jakobovits, 1961). In these tests the subjects are presented with words in two languages and asked to rate the words along scales such as good/bad,
fast/slow, noisy/quiet, pleasant/unpleasant, big/small and heavy/light. These scales correspond to the factors of 'potency', 'evaluation' and 'activity' which are obtained by factor analyses among many languages (Osgood, et al., 1957, etc.) In the former test, when the comparison is made of responses to a word in one language and to its translation in the other language, coordinate bilinguals showed a greater difference between the L₁ word and that of L₂ than in the case of compound subjects. This result was interpreted as indicating a separate system among coordinate bilinguals. The latter test of the "semantic satiation method" is a modified version of the former and it is specially designed to measure the change of judgement as a result of the continued presentation of the same stimuli. For example, the rating of a word 'father' on the 7-point good/bad scale might change after repeated trials from the original judgement made by the subject; [e.g., rating (1) to (7).] This change in estimation as the result of repetition was interpreted as "reactive inhibition" (the term from Hull, 1943). It was reported that a greater amount of reactive inhibition was observed among compound bilinguals than among coordinate bilinguals. It is interesting to note that retroactive inhibition is regarded as the source of compound bilingualism. By implication, it may mean that in the case of compounds, the development of distinctive semantics for L₂ is hindered by the reactive influence from L₁, whereas the coordinate language system is seen in comparison as the more natural outcome of unhindered development. It also appears that for a kind of psychology which advocates an S-R paradigm, the coordinate language system is more readily accepted, since learning languages tends to be viewed as a matter of increasing the number of memory traces which trigger verbal responses and which are highly sensitive to situational cues, (as it is clear from the discussion above (see §8.1)). For these reasons, it would be natural for us to conjecture that the compound language system is viewed by Osgood & Ervin as the unnatural result of converging separate representational mediations, just as the repeated presentation of the same language signs exhibits reactive
inhibition.

The results of these studies were interpreted as illustrating the existence of separate semantic systems among coordinate bilinguals and a single system among compound bilinguals. It is significant to notice that Lambert et al adopted Osgood's semantic differential method and that Lambert & Jakobovits's method is a modified version of the former method. Thus, these studies deal with essentially the same area of semantics. Weinreich (1958) points out that Osgood's semantic differential method elicits only affective meaning. The point is illustrated by the scales used in the semantic differential method (see above). As is well known, Weinreich's criticism was noted by Osgood himself and since then Osgood has regarded his method as dealing with affective meaning. Thus, it appears that the hypothesis of compound vs coordinate language system is closely related with affective meaning. As far as these two experimental studies are concerned, they do not provide as evidence for the separation or singleness of other meanings, particularly conceptual meaning.

As noted earlier, Jakobovits (1970) attributes the dichotomy of compounds and coordinates to attitudinal factors in L2 acquisition. As we have seen in Osgood and Ervin's hypothesis, it is reasonable to assume. Some part of language acquisition reflects a learner's attitude toward the culture of the target language. This interest in a foreign culture may raise his motivation to learn the language. In this sense, affective meaning can serve as an index a learner's motivation, and Jakobovits's view is convincing in this respect. Individually different affective meanings might reflect a varying degree of commitment in language learning as well as the degree of appreciation for the target culture. In this sense it is relevant to consider the hypothesis of compound and coordinate bilinguals in relation to L2 acquisition and learning. Furthermore, it is known that the process of enculturation or acculturation is related to a cognitive
process of accommodation or assimilation (ibid.). These two basic mental activities of accommodation and assimilation are often seen as the principle means by which not only young children but also L₂ learners acquire a language, and by which they interact with other people and the world outside (see Piaget, 1936, 1967, 1975, etc; Corder, 1978b).

§ 8.1.2 Some experimental difficulty involving L₂ subjects

Once we start to deal with bilingual subjects, it becomes difficult to obtain the kind of result usually observed among monolingual subjects involving L₁ materials. Most of the experimental studies discussed below (§8.1.3 & §8.1.4) base their experimental confirmation or disconfirmation concerning the hypothesis of coordinate or compound language system upon previously established psychological experimental results involving L₁ subjects and L₁ materials. Because of the specific difficulties involving L₂ subjects, it is therefore argued below (in the next two subsections) that experimental demonstrations do not necessarily lead to the direct confirmation or disconfirmation of Osgood and Ervin's hypothesis. In this section, we discuss specific difficulties involving L₂ subjects. But at first we try to get some idea of what psychologists call the "automatic processing mechanism" which is informally introduced as preattentive or unconscious reactions below.

The "Stroop Test" is able to illustrate factors that interfere with verbal tasks involving two languages. In Preston and Lambert (1969), Hungarian/English and German/English subjects were asked to name the colour of the ink in which words bearing conflicting values were printed (e.g., the word 'black' was printed in red ink). When the words and the response language were in the same language, the interference was greater than otherwise, i.e., their naming was on the basis of the printed words, not on the colour of the ink which they were asked to name. This appears to suggest that the subjects's immediate response on seeing a word is to read it. In addition
to this kind of preattentive response, most errors appear to be due to other kinds of preattentive responses, since the errors occurred in the items which were visually or phonologically similar in the two languages (e.g., 'blue' and 'bleu').

There have been suggested various models for "automatic processing" which yield preattentive responses: logogen model (e.g., Morton, 1970), levels-of-processing (Craik & Lockhart, 1972, etc), on-line processing (Marslen-Wilson, 1976, Marslen-Wilson & Tyler, 1975, 1980, Marslen-Wilson & Welsh, 1978, etc.) The main purpose of this brief discussion is merely to present a discussion concerning specific difficulties of L₂ subjects contrasted with to monolingual subjects. We will thus refer to only a few points offered by the logogen model, helping us to see more clearly why L₂ subjects pose some experimental difficulties for us.

The logogen model was postulated by Morton (1970); see also Keele (1973). According to Klatzky (1980), the model is experimentally attested by the following well-known studies: Posner & Snyder, 1975ₐ,ₐ; Shiffrin & Schneider, 1977; and Neeley, 1977. This model assumes that various long-term memory codes for an item representing its meaning, sounds and visual forms are associated with a common ground termed "logogen" in which various information from the codes is combined. It is further assumed that neural excitations received from all sources exceeds a threshold value, the logogen makes a recognition decision, determining that a pattern has occurred. Because of these threshold values, the model can account for automatic processing; once the excitations coming from all necessary sources exceed the respective threshold values, it automatically determines that a certain pattern has occurred. While the logogen can register the information activated by attention, attentional activation is limited in capacity. This is distinct from automatic activation; many logogens can be excited at once in the automatic processing mode. Generally, the more familiar to a learner a given item
is, the easier it is processed automatically. In this sense, a learner's familiarity with a given item may be related to the ease of automatic processing; how often a learner is exposed to a pattern may determine how quickly he can recognize the pattern.

It appears that the Stroop test involves the activation of many logogens at once, since it involves the recognition of at least several patterns: graphic forms, colour, sound and meaning. Since the colour of the ink is not a part of the inherent linguistic properties of the words used as experimental stimuli, we may well suppose that the colour of the ink would attract the subjects' attention, particularly when the ink is in contrast with the colour denoted by the words. However, as we have seen above, the task apparently led some subjects to read the printed colour word instead of naming the colour of the ink. This suggests that while automatic processing needs to be established for a learner to be proficient, the experimenter can never tell explicitly how an attentional factor can suppress automatic responses or in what way attention interacts with automatic responses in a given experiment. Besides, in the case of L2 subjects all logogens may not always be equally well developed; some may be as good as those in native speakers, but the others may be far inferior. Also, the threshold values for the respective logogens may be somewhat different from those of native speakers', as we can guess from such cases as l/r phoneme detection among Japanese learners of English (see Dickerson, 1976). Thus, the logogen model suggests that the individually different levels of proficiency create the problem of varied threshold values and the complex interaction of automatic and attentional processing. Unless each of these factors can be estimated prior to an experiment, we can not pin down what the major determinants for a specific experimental outcome are.

Let us take one simple example to illustrate the interaction of attentional and automatic processings. We have seen above that the proficiency level relates to the question
of how much automatic processing is established. Thus, we may suppose that the degree of dominance of one language over the other is proportionate to the ease of automatic processing. In a mixed language experiment (i.e., when bilingual or trilingual lists are used), this straightforward relationship is not reflected in the experimental results, since the less fluent language unexpectedly attracts the subject's attention. Tulving and Colotla's experiment (1970) appears to demonstrate this point. In their experiment, the subjects were presented with monolingual, bilingual and trilingual lists in English, French and Spanish. They were asked to recall the items of the lists in any order (see chap. 6 for free recall). We would expect that the items in the subject's dominant language would be best recalled. The results are contrary to this expectation; although the unilingual lists were best recalled and the trilingual lists worst, the items in the proficient language were mostly forgotten. These results may be regarded as an example of the interaction between attentional and automatic processings for the following reasons.

Among other things, two factors improve the amount of free recall: not only the familiarity of the items to the subject (i.e., subjective familiarity, chap. 6), but also any item the subject attends to is recalled highly well (see chap. 6). The facilitative effect of familiarity in free recall is apparently due to attentional activation of logogens. The attentional facilitation of recall relates to automatic activation of logogens, since one can attribute the successful performance to automatic activation of logogens. In the case of the above experiment, the items in the proficient language were recalled worst of all. When a subject has a fluent command of a language, this would suggest that since he can rely on automatic processing, he does not need to attend to the experimental items so carefully. Whereas, to the items in his less fluent language he needs to attend more carefully. If this is the case, we can guess that, since he attends more carefully to the items in his less fluent language, they can be
recalled better. In this sense, the attentional activation can explain why the recall in the fluent language is impaired most and why the items in the less fluent language were recalled better. But this explanation does not apply to another aspect of the data. As noted above, the unilingual lists were recalled best, bilingual lists next best and the trilingual lists worst. This result may be due to the subjective organization mentioned in chap. 6; i.e., the subjects tend to organize the items of a list to facilitate recall. As we have seen in chap. 6, there are several sources for subjective categories, and the subjective familiarity is one of the important sources. This factor of familiarity would predict that the successful recall due to subjective organization is in the order of unilingual, bilingual and trilingual lists. This predicted rank order corresponds to the observed rank order. It must be born in mind that the factor of familiarity relates to the automatic processing mentioned above. In this sense, it appears that automatic processing is also in operation. Thus, Tulving & Colotla's experiment appears to illustrate the interactive effect of automatic and attentional activation of logogens.

The above issue of proficiency among L2 subjects can be seen in relation to the number of years the subject has studied L2. Then, the general cognitive development is relevant to the issue. As we see below, the cognitive development usually observed among monolingual subjects may not show up in a straightforward fashion. Champagnol's experiment (1973) illustrates the point. The 24-item list were presented to 56 French students of English. The subjects were divided into 4 groups according to the number of years they had studied English. Four different lists were composed: two unilingual lists either in English or in French and two bilingual lists. The items in each list were from four semantic categories. After 12 presentations of each written list, the subjects were asked to recall them in any order in writing. The same test was repeated after one week. The result indicates that, except for the performance of the youngest group the
unilingual lists were better recalled. It is also reported that although translation errors were rare, they increased slightly across the three groups, but negligible for the most advanced group. As we have seen, in a free recall task such as this, the subjects tend to adopt the strategy of subjective organization. This subjective organization is more clearly observable when the items in a list are deliberately organized by an experimenter prior to the experiment (see chap.6). Since Champagnol used pre-categorized lists, the strategy of subjective organization might have been facilitative. This subjective organization is known to improve with age among L\textsubscript{1} subjects (see Koura, 1980). In the case of the bilingual lists this developmental factor was not born out (by this experiment), due to the superior performance in the youngest group. This result, coupled with the increase followed by a sudden drop of translation errors, is difficult to explain. Champagnol postulates the presence of semantic categories which are superordinate to the language specific semantic categories. But without a specific model of memory representation, structure and processing which yields specific experimental predictions, this speculation remains to be seen. Thus, an ideal psychological experiment would require investigations into the recent developments in word- & sentence-processing, memory structures and representations. While we need to study recent models in these areas concerning L\textsubscript{1} subjects, it is also necessary for us to incorporate into experimental rationales the complex issues raised by L\textsubscript{2} subjects as well. Because of these difficulties the present experiment in this section relies on a more direct and conservative method than that taken in psychological experiments (see below).

§ 8.1.3 selective review

This section presents a selective review of experimental studies concerning Osgood & Ervin's hypothesis.

Lambert, Havelka and Crosby (1958), using the "retroactive inhibition" (RI) design\textsuperscript{1}, present evidence that the
distinction between compound and coordinate bilinguals is meaningful and that there is a significant difference between them in terms of the amount of retention they measured. According to the criteria of acquisitional contexts, the subjects were classified into "fused" bilinguals (i.e., compounds) and "separated" bilinguals (i.e., coordinates). The latter group of coordinates was further divided into two: "bicultural" bilinguals who are immersed in two cultures at the same time but who acquired each language in association with these two separate acquisitional contexts (and as a result two separate language systems are formed in their cognition); "unicultural" bilinguals who are immersed in one culture at a time and acquired each language in relation to two different cultural contexts, which results in forming two separate language systems in cognition. The experiment consists of two parts: series I & II. In series I, the subjects learned a list of common English words (trial block 1), followed by a list of 20 nonsense words (trial block 2). In the final phase (trial block 3), the English list was represented and the subjects's retention was measured. In series II the subjects learned a list of 20 English words (trial block 1), followed by a list of French equivalents of the same English words. In the last phase (trial block 3), the subjects's retention was again measured with respect to the same list of English words that was used in trial 1. Normally, if the same item is repeatedly presented to a subject, the item tends to be remembered and it is known that literal repetition usually enhances retention. In the case of compounds, translation equivalents were expected to serve as the literal repetition of their English originals, since translations by definition denote the same mental objects as their originals for compounds; thus, in trial 3, a greater amount of retention was expected among those subjects who were classified as "fused" (i.e., compounds). On the other hand, for coordinates, translation equivalents being mentally represented as distinct from their English originals (since they belong to separate language systems), translation equivalents would not serve as the repetition
of the English originals. In this sense, the experimental design in series II would not be facilitative for coordinates, and their performance would be similar to the results obtained by series I. According to these speculations, the experimenters devised an "independence index", i.e., the difference between the two trials (trial 1 - trial 3) in each series is calculated first; then, the difference between series I & II is calculated:

<table>
<thead>
<tr>
<th>Trial</th>
<th>Trial</th>
<th>Trial</th>
<th>Trial</th>
<th>Trial - Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>E₁</td>
<td>N</td>
<td>E₂</td>
<td>E₁ - E₂</td>
</tr>
<tr>
<td>II</td>
<td>E₁</td>
<td>F</td>
<td>E₂</td>
<td>E₁ - E₂</td>
</tr>
</tbody>
</table>

The experimenters hypothesized that the group of compounds would yield a larger positive "independence indeces" and that the group of coordinates would yield the smaller "independence indeces". This is because series I is not facilitative for a compound but series II is facilitative -- thus, a large positive "independent index" was expected. On the other hand, for a coordinate, series II would work just like series I. Thus, the independent index being the difference between series I & II, a coordinate would yield a smaller and almost zero independence index. The experiment hoped to demonstrate that there was a significant difference between compounds and coordinates in terms of mean independent indeces.

We may well wonder whether a compound would in principle yield a large positive independent index. This question arises in view of what psychologists call "interference by similarity." If a compound does not conceive translations to be equivalent, but to be similar enough to their originals (and it is indeed reasonable to assume so, since affective meanings at least can be different in two languages), "interference by similarity" may be expected even in the case of a compound -- this indicates that Osgood & Ervin's or Lambert et al.'s position, includes other meaning clarity the hypothesis of compounds vs coordinates to although we have confined the hypothesis to the area of affective meanings for the sake of clarity, although
this position is contrary to the discussion presented in § 8.1.1. Once we suppose that "interference by similarity" occurs even in the case of a compound, one can notice that the experimental design series II above would cause inferior retention among compounds. At this point, Young and Navar's experiment (1968) is relevant to the present discussion. Using the same RI design as in Lambert et. al (1958) above, Young & Navar demonstrate that in the case of paired associate learning tasks, "interference by similarity" occurs among their Spanish/English bilinguals. Their subjects learned a list written in one language, followed by a list which contained translations in the other language. They found that associations formed by learning the first list were impaired by "interference by similarity" caused by the translations which were included in the second list. We may not be able to draw a direct analogy from this experiment, but it still appears to be possible to say that presented with translations in the RI experimental design, compounds can suffer from "interference by similarity", which may affect retention under the experimental condition 'series II' mentioned above. Thus, compounds may yield very low independent indeces. If so, the well-known experiment by Lambert et. al (1958) might lose some relevance.

Segalowitz and Lambert (1969) present an experiment which suggests that there is a difference between compounds and coordinates. Their subjects were trained first to generalize concepts from items written in mixed languages. After this training they were asked to press a key when they saw the items belonging to the concepts which they learned to generalize during the first phase of the experiment. The list used for this recognition test contained the originally presented words (i.e., original words condition), synonyms written in the same language that were used during the training (i.e., same-language synonyms condition), synonyms in a language different from the language used during the training (i.e., other-language synonyms) and words which were totally unrelated to the original words. Reaction times (RT) were measured.
The experimenters obtained the following results: original words were responded fastest; other-language synonyms were slower; and same-language synonyms, the slowest. In terms of RT there was no statistically significant difference between compounds and coordinates. However, the experimenters discovered that in comparison with the coordinates, there was a smaller difference between the original words and the two kinds of synonyms among the compounds. The experimenters interpreted this as indicating that coordinates rely on semantic coding to a greater extent than compounds, while compounds tend to pay more attention to visual differences such as spellings. We also note that the experimenters attributed these distinguishing features to the two kinds of bilinguals: coordinates were assumed to possess higher accessibility to semantic coding; compounds were considered to be more sensitive to perceptual differences.

However, Lambert and Rawling's experiment (1969) appears to suggest that an interpretation which is different from Segalowitz and Lambert's might hold. Lambert and Rawling presented their subjects with either monolingual lists or bilingual lists. The subjects were asked to generalize a core concept. The experimenters obtained a result which showed that compounds were better at generalizing the core concept under both of these conditions, i.e., monolingual or bilingual lists. If Segalowitz and Lambert's interpretation above was correct, it would predict that coordinates would show superior performance to compounds', because (according to their interpretation,) coordinates rely on semantic coding more than compounds, implying that coordinates are better at generalizing a concept and recognizing new members of a concept; whereas, compounds attend more to visual differences and can not cope with these tasks efficiently. However, Lambert & Rawling's experiment does not support Segalowitz and Lambert's interpretation of the experiment.

We have seen that Segalowitz & Lambert attributed the two additional features to the two kinds of bilinguals.
Segalowitz & Lambert's view is in a sense an attempt to explain an enigmatic aspect of Osgood & Ervin's hypothesis. In its simplified version (§8.1), the hypothesis cannot explain why coordinates can generalize a concept from a set of words written in mixed languages. This difficulty is the same as the translation task discussed above (§8.1). If there is no link which connects separate language systems, it is difficult to explain why coordinates can deal with any task which requires communication routes between the language systems. Because of this enigma, if we wish to defend the original hypothesis, as in Segalowitz & Lambert, we are forced to attribute irrelevant complications to these two kinds of bilinguals.

Researches into the psychological relation between translations and their originals are relevant as a means of experimental testing of Osgood and Ervin's hypothesis. By using various experimental tasks, this issue has been studied in terms of a question of whether a translation is psychologically equivalent of its original. If the presentation of a translation and its original to a subject serves as a repetition of the same single item, it shows that a compound language system is a more plausible representation. If it does not, it gives a higher plausibility to a coordinate language system as the mental representation of language among bilinguals. For this review, some representative experiments are selected below.

Koler (1968) examines the effect of translations in relation to their originals in a recall task. In a bilingual list the translations used did not contain any visual or phonetic similarity to their originals. Yet, Koler found that both the literal repetition of the same items and the presentation of originals and their translations facilitated recall equally.

Lopez and Young (1974) supports Koler's finding above. In their experiment, the subjects were asked to read a list of common adjectives aloud six times and then to learn a second list which was presented to them aurally. The
second list contained some translations of words which had been on the first list. The experimenters found that the second list facilitated recall.

Evers (1970) also deals with a recall task. German and English words were presented aurally. In some lists translations were included but in others, they were excluded. It was found that the inclusion of translations facilitated recall. On the basis of this result and of a large number of recall errors which were often translations, the experimenter concluded that in spite of the aural presentation bilinguals remember words in terms of their underlying concepts and that they have difficulties in noting which language was used, even when the subjects were specifically instructed to do so.

Kintsch (1970) also found that translation had the same effect as the mere repetition of the same item in a recognition test. After learning a series of bilingual lists the subjects were given a recognition test. Group I were asked to judge whether each item was 'new', but repeated in the 'same' language, or repeated in the 'other' language. Group II were instructed to answer 'yes' only when an item was repeated in the same language. Group III were instructed to say 'yes' when an item was repeated in the same language as well as when an item was a translation in either of the two languages. In Group III there was only a slight difference between simple recurrence of the same items and the presentation of translations (95% vs 87%): Kintsch (1970:407). In Group II there was a high percentage of false recognition, suggesting a strong tendency among the subjects to regard translations as the same as originals. The results appear to suggest the presence of a unified semantic code across languages.

Kintsch and Kintsch (1969) also examines the effect of translation upon memory experiments. Kintsch & Kintsch used a paired-associate learning task. The control group learned a list of 4 English adjectives and 4 German adjectives which were paired with digits 1-8.
experimental group learned a list of 4 English adjectives and 4 German translations of the English adjectives each of which were paired with numbers 1 through 8 as in the control group. In paired-associate learning among monolinguals, the similarity among the items of a list is known to cause interference (Wickelgren, 1979). This "interference by similarity" tends to either impair the memory storage which leads to forgetting, or to create 'response competition', which leads to erroneous responses (ibid.) In Kintsch & Kintsch (1969) this interference through similarity was observed in the experimental group. The result was interpreted by the experimenters as suggesting that the subjects rely on semantic coding rather than visual or phonological coding and that words denoting the same state in the real world are conceived by the subjects to be similar enough irrespective of the difference of languages.

Lopetz et. al(1974) supports the result above in the same paired-associate learning task, using a retroactive inhibition design (see footnote 1). In the retroactive inhibition design if the items in the two lists presented to the subjects consecutively are similar, we normally obtain some interference effects from the second list. In Lopetz et. al four combinations of lists were presented aurally to the subjects followed by four kinds of second lists: (1) only the language of the paired items was changed; (2) items were changed but the language was the same as in the first list; (3) both items and language were changed; (4) entirely novel pairs were presented. The experimenters found that erroneous responses were caused by translations more frequently in conditions (1) & (3). This result appears to suggest "involuntary processing at the semantic level with apparent disregard" for the difference of language: Albert & Obler (1978:68).

As noted earlier (chap. 6), in free recall, words belonging to the same semantic category tend to be recalled as a group. Lambert et. al (1968) examined whether this category effect, which has been recognized among monolingual
subjects, can occur among bilingual subjects as well. If the bilingual subjects can categorize items written in mixed languages, the compound language system appears to be more plausible as the representation of their internal language system. In Lambert et. al, 24 French/English bilinguals and 24 Russian/English bilinguals were exposed to 10 presentations of 40-item lists. In the first two and last two presentations, the lists were monolingual, involving unrelated words. In the third and the tenth presentations, semantically unrelated words were presented in mixed languages. This was to avoid both 'recency' and 'primacy' effects. During the four presentations in the mid-course of experiment, the experimenters used semantically organized lists; their organization was based on four semantic categories in the two languages. In the "concordant" lists, the items were presented consistently in one language and all of them belonged to the two semantic categories. In the "discordant" lists the items were written randomly in either of the two languages, but the category concepts could be intuited by the subjects if the subjects disregarded the difference of language. The experimenters found that the "concordant" lists were better recalled, since the "discordant" lists produced the highest number of translation errors. The result is interpreted by the experimenters as indicating that, while the consistency of language has an obvious psychological effect, cross-language categorization clearly exists, although cross-language categories may not be as cohesive as monolingual categories.

Goggin & Wicken (1971) supports the interpretation offered by Lambert et. al (1968). Goggin & Wicken tested whether a shift in language was functionally equivalent to a switch in semantic category. Their experiment appears to be relevant as a means of experimental testing of Osgood & Ervin's hypothesis in that if a shift in language functions as a shift in semantic category, it is more likely that a language as a whole forms one single superset (category) for all the linguistic expressions contained in a given language. If so, we may take this experimental testing
as offering possible evidence for the coordinate language system, because each language is to be represented as forming one separate category. However, an alternative explanation is also possible; even if a shift in language improves a subject's memory performance, just as a shift in semantic category does, it does not necessarily mean that all the linguistic expressions belonging to one language are categorized into one group in the mind of a subject. This is because what psychologists call "release from proactive inhibition" can explain the improved memory performance. Keppel & Underwood (1962) demonstrated that performance declines from trial to trial as "proactive inhibition" (i.e., detrimental effects of previously learned items on the subsequent learning) develops. In order for an item to be retrieved, the item must be discriminated from the other items which have been presented previously. This discrimination becomes more complex, as there are more previous items. By making items in a current trial easily distinguishable from the previous items, experimenters can improve a subject's memory performance. This phenomenon has been called "release from proactive inhibition". This release is experimentally induced by a shift in modality (e.g., consonants to digits) or a switch in semantic category: Wickens et al., 1963; Wickens, 1972, etc. Goggin & Wicken found that changing both language and semantic category yielded the highest increase in recall; the change of semantic categories (but not the language), the 2nd highest; the change of language, the least increase in recall. For this result, one can draw a similar conclusion which is presented by Lambert et al (1968). It is undeniable that a change in language can have some psychological effects, but a shift in semantic category has stronger effects, which may have some indirect bearing on cross-language categorization mentioned above.
§ 8.2.0 experiment

We have seen above that Osgood & Ervin's hypothesis is difficult either to confirm or to disprove experimentally (§8.1.2) and that as far as the experiments reviewed here are concerned, the issue still remains inconclusive (§8.1.3). Yet, the hypothesis is relevant to the present research in that if the coordinate language system is plausible, the idea of translatability based upon a learner's native language does not follow automatically. Assuming the hypothesis is relevant to the present research, we cannot leave the issue unsettled. The previous experimental testings which have yielded ambivalent results were indirect in their approach in that they investigated the way in which a single or double language system is reflected in a subject's memory performance. Here we adopt the more direct method of eliciting a subject's assessment of two semantic features of three or four sentences in English and Japanese. Regarding these two semantic features, we compare English with Japanese. If there is a difference between them, it would suggest a coordinate language system. If there is not any significant difference, it would support a compound language system. If the subject's assessments support a coordinate language system, we have to make some necessary provisions to accommodate the idea of translatability into the evaluative framework.

The semantic features chosen below are those which were analized by MDS as the bases from which the subjects derive their sense of similarity between the paired NL sentences (see chap. 4, 9 & 10). As regards to these semantic features, we will examine whether the subjects' assessments favour two separate language systems, one for NL and the other for TL, or a single language system common to the two languages.

§ 8.2.1 administration

42 2nd year university students took part in the experiment.
They were presented with 3 kowasu (break) sentences, 4 yaburu (tear) sentences, 4 kudaku (smash) sentences and 4 oru (bend) sentences. They were asked to rate these sentences on the 7-point scales of concrete/abstract and basic/non-basic. Although Kellerman called one of the dimension "core/non-core", this label appears to be difficult for the subjects to comprehend. For this reason, it is called here 'basic/non-basic'. The order of presentation was counterbalanced. One hour later, they were presented with English translations of the above 15 sentences and asked to rate them using the same scales. Half of the subjects followed the reverse order, that is, 15 English sentences first, followed by the 15 Japanese sentences. 7 subjects who did not know the English words were excluded from the analysis.

§ 8.2.2 result & discussion

The data is analyzed in three ways: (1) the Wilcoxon matched-pairs signed-ranks test; (2) the t-test; (3) the probability ellipses.

(1) On the assumption that human judgements are basically ordinal rather than interval (see chap. 5), the Wilcoxon test is useful for the present purpose, since we are interested in finding out whether there is a significant difference between English and Japanese with respect to the two semantic features relative to the 15 expressions. This Wilcoxon test serves as the t-test in the case of ordinal data: Siegel(1956:83). While the t-test is useful for the interval data, the Wilcoxon test can compare the subject's ordinal rating of a Japanese sentence with his ordinal rating of the English counterpart. As Tables 1 - 4, 5.2, Appendices show, none of the 15 Japanese sentences differ significantly from the 15 English counterparts in terms of the two semantic features. Thus, in terms of the Wilcoxon test, the compound language system is supported by the data.

(2) Although the ordinal assumption is more appropriate
than the interval assumption in view of the intrinsic nature of human judgements (see chap. 5), the data is analysed by the t-test, just in case the subjects' metalinguistic judgements are reliable enough to warrant the interval assumption. As Tables 1 - 4, 5.2, Appendices show, 5 out of the 30 judgements do not satisfy the assumption of equal variance. Of the 25 comparisons, 2 cases yield a significant difference: i.e., on the scale of abstract/concrete, *dan-boru bako wo yaburu* (break/tear a cardboard box) and *koigokoro wo yaburu* (break one's heart).

(3) In the t-test above, the mean ratings of the English sentences are compared with those of the Japanese sentences with respect to each of the two semantic features separately. The probability ellipses can represent pictorially how much the mean ratings of the English sentences with their variances and correlation coefficients differ from those of the Japanese sentences with respect to the two semantic features in combination. As we have seen in chap. 7, we assume that the subject's judgement in response to each semantic feature (or a bundle of features as a whole) forms a normal distribution. If this assumption is tenable, the joint distribution of the two semantic features takes in theory a form of bivariate normal distribution function. If we slice the bivariate normal density function so that the incision contains the two means along the two axies (i.e., the two semantic features), the resultant pictorial representation reconstructs how the subject's judgement with respect to the two semantic features of Japanese differs from that of English. Fig. 4 offers some general guidelines to interpret the probability ellipses.
Fig. 1 - 19, in 5.4, Appendices are the probability ellipses computed by the PPSS program. The representations are in keeping with the t-test results, since this method of probability ellipses also adopts the interval assumption.

In the case of the t-test we have seen that the three pairs of Japanese and English sentences (i.e., kowasu 1, kudaku 1 & oru 1) did not satisfy the assumption of equal variance. The probability ellipses for these pairs also show the strikingly different variances between the corresponding pairs. One feature common to these pairs is that the Japanese ellipses are contained in the areas of the English ellipses. This suggests two points.

(1) These three pairs do not support the coordinate language system. It may be recalled that of 30 pairs, the t-tests showed no significant differences for the 25 pairs, favouring the compound language system. We have now seen that 3 out of 5 pairs which the t-test could not be applied to favour the compound rather than coordinate language system.

(2) The present research is concerned with the role of analogical reasoning in L2 acquisition. This concern is operationally translated into the problem of a learner's translatability judgement based upon the NL similarity. In order for this operational translation to be plausible, the learner-language has to favour the compound rather than coordinate language system, as we have argued previously. So far, of the 30, we have seen that 28 pairs do not pose any theoretical difficulty in this respect.

The t-test shows that yaburu 1 & 4 involving the abstract/concrete feature are significantly different from the English counterparts. The probability ellipses for these sentences show that while the Japanese ellipses considerably overlap with the respective English ellipses, there is a smaller disjoint area between them. Due to the respective overlaps, these sentences do not necessarily support the coordinate language system, favouring again the compound language system. It is also likely that a
learner's 'sense of translatability' can be derived from these overlapping areas of NL and TL. Thus, it appears to be reasonable to assume that, starting with the conceptual equivalence between a learner's 'sure' TL expression and his NL counterpart, a learner's judgement of translatability can be extended to other NL items, based upon the NL similarity between the 'sure' NL item and the other related NL items.

It may be recalled that the two semantic features considered here are those which the MDS analysed with respect to the Dutch sentences in Kellerman's study (1977). As far as these semantic features are concerned, it appears to be difficult to endorse the coordinate language system in terms of a learner's metalinguistic judgements in the present experiment. Unless one can elicit responses which warrant the cognitive separation of TL and NL, one can not demonstrate the plausibility of separate language systems. However, even if we manage to obtain responses which are sufficiently different from each other, we will have an equally difficult task of whether or not they are in fact conceived by a learner to be identical, similar or different. However, as we have seen, the present experiment is not specifically concerned with determining which of the two possible language systems can better account for our learner-language. We are concerned with the minor point of whether our basic rationale for the present research is plausible or not, i.e., all learners participating in the experiment are capable of judging translatability, assuming that they do not possess the extreme form of coordinate language system (see §8.0 & §8.1). In this respect, we have seen that the assumption is reasonable relative to the data analysed here.

Diller (1974) points out that no two languages are similar enough to be "compoundable" nor dissimilar enough to be coordinate. This statement appears to be reasonable in view of the probability ellipses for the English and Japanese sentences, because the Japanese ellipses are not completely identical to the English counterparts, nor
dissimilar enough to justify separate or disjoint language systems, one for Japanese and the other for English.

§ 8.2.3 conclusion

We have seen in §8.1.1 that the hypothesis of compound vs coordinate language system is more closely related with "affective" meaning than the other kinds of meaning. However, the hypothesis has been presented as involving more general aspects of languages. This suggests, as it has been argued, that the hypothesis involves proper considerations of memory representation, structure, encoding and decoding processes (§8.1 and §8.1.2). These are the issues which can be treated properly by psycholinguistic research methods. However, as we have seen in 8.1.2, psycholinguistic experimentations involving L2 subjects, particularly in the case of experimental testing of this hypothesis, become very difficult to run. Unless we can explicitly formulate the structure of memory and representation as well as the encoding and decoding processes, we can not be sure what causes variable performance on the surface. Probably because of these indeterminable complex issues, psycholinguistic investigations into this hypothesis have been inconclusive (§8.1.3). We have adopted a more direct method of eliciting judgements in response to the NL and TL. This means that the present experiment does not have direct bearing on the hypothesis. The present experiment is merely concerned with the immediate question of whether the main purpose of the present research (whether NL can play a role in learning L2, with respect to judgements of translatability as derived from the general mental activity of analogical reasoning) is plausible. Since the hypothesis of a coordinate language system as it makes the task of translatability judgements theoretically impossible (§8.0, §8.1, & §8.2), (it was necessary to test this.)

The experiment dealt with 3 kowasu sentences, 4 yaburu sentences, 4 kudaku sentences and 4 oru sentences. The
subjects were asked to rate these sentences on 7-point scales of concrete/abstract and basic/non-basic features, both in English and in Japanese. The rated scores are analysed in three ways to find out whether the ratings in the NL are significantly different from those in English.

(1) The Wilcoxon test regards the data as ordinal. The result shows that none of the 30 pairs are significantly different.

(2) The t-test regards the data as interval and requires the respective variance to be 'equal' in terms of the F-ratio. 25 pairs satisfy the latter condition. Of these 25, the two pairs (yaburu 1 & 4) with respect to the concrete/abstract feature show significant differences between Japanese and English.

(3) However, in view of the probability ellipses, the 5 pairs which did not satisfy the assumption of 'equal' variance were shown by the PPSS program that the Japanese ellipses were contained in the areas for the English counterparts. Further, the 2 pairs which showed significant differences according to the t-tests involved considerable overlapping areas, having a smaller amount of disjoint areas between the paired ellipses. Moreover, the other Japanese ellipses for which the t-tests showed no significant differences overlap considerably more with the English ellipses than in the case of the two pairs of ellipses we have just considered (i.e., those for which the t-tests showed significant differences between the NL and TL). Since it is possible for a learner to derive his judgement of translatability from the overlapping areas between the TL and NL, it is at least reasonable to assume that the present research into a learner's judgements about translatability based on his NL is justifiable, though the hypothesis of coordinate language system makes the translatability judgement difficult to perform.
Chapter 9 Pilot study (6): experiments to determine a form of processing

§ 9.0 purpose

The follow-up study in chapter 4 incorporated the hypothesis of processings (integral vs separable) into the analysis of data. However, the study did not contain any separate experiment which could determine the specific form of processings the subjects adopted at the time of their making similarity judgements. For this reason, the analysis involved some unnecessarily speculative aspects and yielded more than one analysis to account for the data. It would be more desirable to obtain a single analysis. Thus, chapter 4 concluded that it would be more fruitful if we could run an independent experiment which would determine a specific form of processing. The purpose of this chapter is to present some such experiments.

The same five groups of subjects who took part in the translatability experiments in chapter 10 participated in the present experiments. Since the subjects were grouped according to their school years, an additional interest of this chapter is to see whether the subjects adopt different forms of processings according to their ages. Although school years may not correspond to their cognitive development directly, they may offer some indirect measures of cognitive development. Thus, we might be able to make some suggestive remarks concerning the relationship between processings and cognitive development. But, the main purpose of this chapter is to determine a specific form of processing according to the 5 groups of subjects. We intend to prepare a more focussed analysis in chapter 10 so that the subjects' judgements of translatability can be analysed in relation to the ways in which the subjects process the similarity relationships among NL items.

§9.1 gives an overview that MDS can be used for the purpose of determining a form of processing. §9.2 considers 6
types of processing: integral, separable, asymmetric separable, 'configural', 'dominant' and 'categorical'. The section also presents an overview on the relationship between the 6 types of processing and their underlying metrics. §9.3 describes the experiments: §9.3.1 presents the experimental method which is derived from §9.1 and §9.2. §9.3.2 discusses the results obtained. §9.4 offers a tentative conclusion.

Garner's definitions of integral, separable, and asymmetric separable processings are illustrated in chap. 4, but his own definitions are quoted below. Both the illustrative account of Garner's view in chap. 4 and his proper definitions in this chapter are necessary to this thesis. The actual experiments in chap. 4 require the former and the more theoretical issue of processing in this chapter requires the latter. As for the configural processing, Garner's definition is presented in §9.2. The present discussion of dominant processing is based upon Arnold(1971), Goldman & Homa(1977) and Shepard(1974). The useful notion of unit circle is also introduced in §9.2. This is because the shape of the configuration in the MDS result can conform to a unit circle and the distinct unit circle determines a specific metric (hence, the form of processing adopted by the subjects).

The idea of the categorical processing is derived from Shepard(1974). According to Shepard, linguistic items are not particularly suitable for the dimensional scaling adopted by the MDS. This is mainly because semantic features are not strictly dimensions (see below). Semantic features are rather categorical or discrete. The type of processing which is suitable for these categorical or discrete mental entities is called here categorical. Shepard's definition of his non-dimensional representation is adopted below. We characterise Shepard's definition in relation to the theory of lattices or Boolean algebra. We intend to relate Shepard's definition to Rosch's concept of word-meaning (chap. 3). By doing so, we identify the categorical structure as equivalent to either of the
two paired expressions (either eq. 3 & 4 or eq. 5 & 6 below). In the course of discussion, we pave our way toward connecting the present application of Rosch's and Shepard's definitions with Piaget's accommodation and assimilation. Consequently, the IL notions of recreation and restructuring processes of TL acquisition (see chap. 2) will be experimentally demonstrated in view of Piaget's framework in chap. 10. Then, the experiments of analogical reasoning in chap. 10 (which is interpreted here as the subject's reliance on the NL similarity in making a translatability judgement) can be seen more properly as an experimental method which delves into the acquisitional process involving accommodation/assimilation in correspondance with recreation/restructurization in chap. 10.

The lengthy discussion to characterize the categorical processing in §9.2 is relevant for the purpose of this chapter. There have not been suggested any specific metrics in correspondence with this processing. However, the characteristic features of categorical processing will clarify that there is one salient, overt behaviour which can identify the processing, if it should actually be in operation in the judgemental data we elicit. But this salient feature does not become apparent to us until some logical characterization is available to us (see below). For this reason, the categorical structure/processing is discussed in detail below.

§ 9.1 another use of MDS

§5.3.1 characterizes the method of MDS as a subject-oriented exploratory procedure which is meant to discover some underlying criteria for the similarity judgements the subjects make. As input data MDS requires only the judged similarities of all the possible combinations of paired stimuli; and an outside investigator's intervention may be minimal. Since MDS can abstract such tacit dimensions that are not contradictory to the ordinal (or interval
information involving all these combinations, we have seen that it is reasonable to regard MDS as subject-oriented procedure. Since IL studies aim to be learner-oriented, MDS suits this purpose of IL studies (see chap. 4 & 5).

MDS is exploratory in the following sense. It produces a map in which a given set of stimuli is plotted against the conceptually 'interpretable' dimensions, retaining at the same time the ordinal or interval information supplied as input. Because this configuration of stimulus points can fit excellently to input data, and the derived dimensions can be interpreted (i.e., dimensions are conceptually understandable to an outside investigator), the originators have claimed that the MDS representation might resemble or reflect subjective states in cognition (e.g., Shepard, 1972). In this sense, MDS is exploratory, delving into cognition. When experimental materials are semantic, the above claim goes even further; MDS aims to reconstruct some semantic states stored as a part of long-term memory (e.g., Rumelhart, 1977).

There is substantial experimental evidence which can endorse the above cognitive claim: Rumelhart & Abrahamson, 1973; Rips, Shoben & Smith, 1973; Caramazza, Hersh & Torgerson, 1976; Shoben, 1976; Homa & Omohundro, 1977, etc. These experiments demonstrate that semantic distances computed by MDS can predict various types of inference processes fairly accurately. For this reason, we can at least indicate the following. Since the method of MDS can 'reconstruct' a subjective space in which the derived axes reflect implicit criteria of the subjects' similarity judgements and each stimulus point is 'placed' relative to each criterial axis and to each other, the derived configuration is a possible candidate for a mental representation in cognition, although an MDS representation may not be a 'genuine' semantic map in cognition. We may also recall that the free recall experiments in chapter 6 also present some evidence for this cognitive claim; we have obtained fair correspondences between the clusters.
in the MDS configuration and those in free recall.

We have seen above that the representations obtained by MDS are justified to some extent for its cognitive claim by some experimental evidence. The claim gains more credulity when we recall that the MDS computation includes a procedure (called MINKOWSKI in MINISSA program) which allows for various options for possible processings the subjects may adopt in making similarity judgements. That is, an experimenter can adjust an appropriate parameter for his assumed underlying metrics in correspondence with a specific form of processing which he assumes that his subjects have adopted. The best selection of parameter would yield the smallest 'stress' value (i.e., the index representing the discrepancy between input and output).

Thus, by looking at various stress values which the respective parameters yield, one can conclude that the metrics which yields the lowest stress value of all denotes a specific type of processings the subjects have adopted. At this point, it is important to note that this parameter setting is not the only factor which determines a stress value. The question of dimensionality (i.e., the relevant number of dimensions) also contributes to a stress value. Generally, the more dimensions there are, the smaller its stress value becomes. Yet, we wish to obtain a smallest set of 'criterial' dimensions which are interpretable. Thus, the present issue requires a delicate balance between dimensionality and underlying metrics.

As asymmetric separable processing illustrates (see next section), MDS does not always reflect the inherent structure of input stimuli, but it is possible to reflect performance factors. This is because the subject chooses a specific processing in order to accomplish a given experimental task effectively. It is important to bear in mind that the MDS representation is the product of the subject's judgemental performance about the extent of similarity of the paired items.
As we review in the next section, there is an orderly relationship between processing and underlying metrics, but the relationship is not deterministic or biunique: e.g., city-block metrics refers to either separable processing or asymmetric separable processing. Furthermore, the relationship between processing and an inherent structure of stimuli is not deterministic (see, Kemler & Smith, 1979). As Foard & Nelson (1984) and Smith & Nelson (1984) make clear, the structure of stimulus or a bundle of specific properties is merely one of the three major factors whose complex interaction appears to cause the subject to adopt a specific form of processing. These factors are (1) the nature of experimental task, (2) the level of cognitive maturation and (3) the nature of stimulus structure (ibid.) When the stimulus structure is inherently 'separable' or 'integral', the subject exhibits a tendency to adopt a separable or integral processing respectively, but it is not necessarily so (see Sepp, 1978). Thus, it appears to be safe to suggest that the subject is capable of manipulating different kinds of processing, depending upon a specific demand of an experimental task and the subject's level of cognitive maturation. For this reason, we cannot expect that the MDS can reveal the structure of stimuli in a straightforward manner, since the similarity judgements can be seen as the result of these factors as well, i.e., they are task-specific, and cognitively governed products.

The special relevance of the discussion above is that the subject's adoption of specific processing is the outcome of various performance factors including the factor of a specific nature of experimental task, such as the requirement of speedy processing which may cause the subject to pay attention to the most relevant features and to ignore the rest of the features stored as a part of his knowledge of language. At this point, it is relevant to recall our informal distinction between accessibility and availability of our knowledge of language in §5.1. We have noted that competence (one's entire knowledge of language) is constantly accessible, but it is not always available to a language user (see chap. 3, §3.4 and chap.
We have seen that the subject's adoption of specific processing is derived from not only the structure of stimuli itself (i.e., in the case of linguistic materials the structure represents the subject's structural understanding). But also the subject's choice, conscious or unconscious, is determined by various performance factors. The same holds true in the process of the subject arriving at similarity judgements which the experimenter uses as the input to MDS. In order for the subject to derive his similarity judgements, his competence (i.e., mental representation about his structural understanding etc.) will be consulted. However, depending upon the nature of the experimental task in relation to the linguistic materials used, it may not be necessary for him to scan all the knowledge he has exhaustively. Rough mental examination of a few relevant features may be sufficient. In this sense, although all linguistic competence is accessible to him in principle, all knowledge does not have to be available to him, in practice. The extent, manner and place of availability appears to be determined by his specific form of processing (see below), and the choice is dictated by the three major interactive factors mentioned above: (1) the nature of task relative to materials used; (2) the level of cognitive maturation; (3) the current state of the subject's growing competence as his source of information. For this reason, it is highly meaningful that the MDS computational procedure takes into account the possible influences derived from a specific selection of processing in terms of its underlying metrics. The foregoing discussion concerning the MDS computational device may be summarized by the following diagram. It involves three components. Although these are represented sequentially, it is merely for the sake of convenience.
Diagram 1
The upper arrow (A) represents the route through which the subject arrives at the judgement of similarity. The lower arrow (B) represents the computational route followed by MDS; starting with the reverse direction to the subject's judgemental process, MDS reconstructs a mental map. In (A), consulting his personal semantic map in cognition, the subject adopts an appropriate form of processing when he is exposed to an experimental task. Then, he draws his conclusion about a particular pair relative to the other pairs. In (B), MDS operates backwards. Starting with the input data of similarity judgements, MDS loads an experimenter's assumption about his subject's specific form of processing in terms of its underlying metrics (Euclid, city-block, supremum, etc. below). Then, MDS computes the configuration of stimulus points. The outcome is the MDS version of reconstructed 'semantic map' for the subject. Since the experimenter can alter the parameters for his assumed underlying metrics, he can compare respective representations according to each of these parameters. He can also obtain stress values representing how much the reconstruction diverges from the input. The lowest stress value along with the problem of dimensionality and interpretability mentioned above suggests the best goodness-of-fit to the input and if the input we obtain experimentally were not strongly biased by performance factors, we might regard the MDS reconstruction as a good candidate representing some semantic state in cognition.
We have seen above that we can make use of MDS to determine a specific form of processing the subjects adopt at the time of making their similarity judgements. Because MDS enables an experimenter to manipulate various kinds of metrics, he can see which underlying metrics yields the best goodness-of-fit to the data. The present procedure is legitimate; as we see below, several psychologists have argued and experimentally proved that there is a definite correspondence between metrics and processings. Following this procedure, we can understand the importance of a reason why MDS provides the option of parameter-setting with respect to underlying metrics. The availability of this option in MDS appears to contribute to its cognitive claim we have discussed above.

§ 9.2 the relationship between processings and metrics

Garner's review article (1976) draws our attention to the fact that there is a definite relationship between processings and metrics. But the relationship itself is made note of by Torgerson as early as 1958 (see Torgerson, 1958, chap. 11, particularly p. 254 where he relates city-block metrics to processing). Garner (1976) considers 4 kinds of processing: integral, separable, asymmetric separable and "configural". Garner relates the first three types of processing to the following metrics: Euclid, city-block and city-block. Garner does not specify the metrics for what he calls 'configural'. But, Garner explains this processing in terms of visual properties. Since we are essentially concerned with conceptual rather than visual properties, the absence of explicit specification of metrics for configural properties does not concern us much here.

Garner's view on integral and separable/asymmetric separable processings on the basis of Shepp (1978). At this point Garner's definitions themselves are introduced to clarify the illustrative account above. But first of all Garner's configural processing
is presented. Although we cannot specify the underlying metric for this processing, a brief discussion may be relevant in the present study, since the configural property refers to the objects whose geometric attributes are important. Some of the object-nouns in the present study include these geometric properties.

According to Garner (1974a, b, 1976), the configural structure of a stimulus represents "the case two dimensions combine to form a qualitatively new dimension"; Garner, (1974a:164); for the definition of dimension, see below. Garner gives an example of "area" where width and height are metric dimensions. When these dimensions are combined, a new dimension of "area" is in effect produced. Garner also explains the configural processing in terms of (1) the impossibility of selective attention to the relevant dimensions and (2) of no facilitation upon processing, when the dimensions are correlated. The first point means that in processing something called "area", the two notions of width and height are processed in conjunction to each other. The second point means that because the concept of "area" is an entity which needs to be grasped wholistically, the correlated dimensions do not facilitate processing -- Garner's use of the term "correlated dimensions" is illustrated in chap. 4 and below.

In Garner's explanation of integral, separable/asymmetric separable processings there are three basic terms: "dimension", "correlated" and "orthogonal". These are presented first, since they are used, not only to characterise the structure of both a stimulus itself and the set of stimuli, but also to define the three types of processings he considers.

Garner appears to use the term dimension in two related meanings. The basic meaning of 'dimension' represents some property which is continuous such as the colour spectrograph, the electrograph of sounds. The continuum for sounds or colour have components which can be represented by further constituents which are also
continuous: eg., amplitude, the shape of a sound wave, brightness, saturation etc. But the term is also used to refer to discrete features. Garner writes: "a traditional term 'feature' is equivalent to the specific values on a dimension", Garner, 1974a:6. The 'apparent' value on a underlying continuous dimension makes a specific feature appear to look as if it were 'discrete'. But it is necessary to assume that this 'discrete' value is a part of and is derived from some such underlying continua.

Garner's own example illustrates the above point (Garner, 1976a:100). The set of stimuli consists of 4 members: a red circle, a red square, a green square and a green circle:

![Diagram of color and form dimensions](image)

In presenting this set of stimuli, Garner assumes that there are two dimensions of colour whose polar extremes are red and green, and of form whose polar extremes are circle and square (see fig. 1). At first glance, we might assume the four members of the set to be 'discrete'. However, Garner regards the two dimensions as continuous. This interpretation is justified by assuming that in the case of the form one can think of an origin at which the two dimensions (of polar features of square and circle) meet; that is, the origin is the smallest form or area which is neither square nor circle, approaching infinitely to a point. This origin stands at the intersection of the other dimension of colour. The same argument applies to this dimension of colour. As in the dimension of form, one can think of an origin which is neither green nor red. In this case, the polar points are the brightest green and red. Then, one can
recognise that this dimension is also continuous. Because the idea of origin is plausible in both dimensions without destroying the independence of the two dimensions, Garner's use of the term 'dimension' assumes the underlying continuous quantity or quality. Thus, the set of four members (a red circle, a red square, a green square and a green circle) are values derived from the two continuous dimensions.

It may be noted incidentally that 'circle' is not usually regarded as the polar feature of 'square'. However, the assumption of some 'origin' which intervenes with circle and square enables us to postulate the continuous dimension. Relative to this framework, 'circle' becomes operationally polar to 'square'. The same argument appears to hold in the case of the 'colour' dimension.

In the above, we have seen that Garner's use of the term dimension involves two meanings: (1) the usual notion of continuous quantity or quality (the notion applies to the latter only if it can be converted into some quantitative representations); (2) the discrete value derived from the continuous underlying quantity. We have stressed that Garner's terminology basically assumes some continua. Now we will proceed to see how 'orthogonal' and 'correlated' dimensions are defined by Garner. We have already illustrated these in chap. 4 and here we will define them again in relation to the above interpretation of his continuous dimensions.

We may note first of all that, because dimensions are understood to be continuous, some numerical representation of stimulus points are made possible, i.e., one can consider the set of stimuli in a multi-dimensional space, such as seen in MDS. Garner uses the term orthogonal in reference to the uncorrelated dimensions. This means in our example of fig. 1 that the dimension 'colour' is independent of another dimension 'form'. Thus, the set of four members (a red circle, a red square, a green circle and a green square) forms an orthogonal set. Whereas,
a smaller set of two members (a red square and a red circle) is correlated set, since the dimension of colour is kept constant and the values along the dimension 'form' is made to vary. This distinction of orthogonal or correlated set of stimuli is as important as Garner's definition of dimension, since Garner differentiates his 4 types of processings according to his distinction between orthogonal and correlated set.

Garner defines his 4 types of processings in terms of facilitation or non-facilitation of discrimination and whether selective attention is possible or not. The following table summarises Garner's view:

(1) integral: "selective attention to the dimension is not possible with orthogonal dimensions"; 1976a:100 "facilitation occurs when the dimensions are redundant"; ie., correlated; idem.
(2) separable: "selective attention is possible with orthogonal dimensions"; 1976a:101; "no facilitation of discrimination with correlated dimensions"; idem.
(3) asymmetric separable: "facilitation in discrimination occurs with correlated dimensions; 1976a:102; "selective attention is possible with orthogonal dimensions"; idem.
(4) configural: "selective attention to the dimensions is not possible"; 1976a:101; "no facilitation when they are correlated"; idem.

Table 1

Arnold (1971) considers the following kind of processing. Arnold explains it as a process in which the subject "suppresses all dimensions except the one that maximally discriminates the members of a concept pair"; Arnold, (1971:349). That is, the subject attends to one specific 'salient' dimension, ignoring the rest of the dimensions. Arnold regards this processing as related to maximum (1971:355-7) and he expresses it in terms of Minkowski's general metrics as follows:
According to (1), Arnold's processing is equivalent to what Shepard (1974) calls "supremum metrics" (1974:405) and to what Goldman and Homa (1977) term "dominance metrics"; Goldman & Homa, (1977:377, footnote 1). Goldman & Homa describes their processing as follows:

"The dominance metric essentially weighs most heavily that dimension which maximally separates two points (stimuli). For the present study, the dominance metrics was defined with r = 50." (ibid.)

The above citation shows that the "dominance" metrics represents the same mental process that is proposed by Arnold. Shepard's terminology of "supremum" metrics may suggest that r= . But, there must be an upper bound before r reaches ; and in practice, supremum usually represents maximum. Shepard(ibid.) indeed expresses the supremum metrics as

\[
(2) \quad d_{jk} = \max_m |a_{jm} - a_{km}|
\]

Since it is not possible to compute even maximum in reality, both Arnold and Goldman & Homa use as a parameter of r what they consider to be a fair approximation; in the case of Arnold, r = 32.0; in Goldman & Homa, r = 50. This means in our situation r = 12.0, under the current restriction of the MINISSA program we use.

In the above, we have seen that Shepard's supremum metrics, Golman & Homa's dominance metrics and Arnold's metrics for the "maximum component" spaces (op.cit.) denote the equivalent expression. This processing is not given any special name by these researchers. Here we will call it 'dominant processing' and adopt Arnold's definition cited above; i.e., the 'dominant' processing represents a cognitive process in which the subject attends to one
compelling dimension, suppressing the other criterial dimensions which may be stored as a part of his linguistic knowledge.

We notice that there is a strong affinity between Garner's definition of 'asymmetric separable' processing and our definition of dominant processing. In both cases, the subject attends to one obvious and compelling dimension.

In spite of the apparent conceptual similarity between these processings, Garner assumes that the asymmetric separable processing involves the city-block metric \( r=1.0 \), while according to the other researchers above, the dominant processing calls for the distinct supremum metric. It is counterintuitive to see that the two notionally similarly defined processings require different metrics. This discrepancy makes us wonder if the hypothesis of 'asymmetric separable' processing may be redundant. In the case of the dominant processing, the conceptual definition above is mathematically realized, but it is not so in the case of asymmetric separable processing; ie., in terms of metrics, the asymmetric separable processing is the same as the separable processing and both are equally represented by the city-block metrics. Furthermore, in the dominant processing, a large value of \( r \) would yield a proportionally smaller stress value. But since Garner's notional distinction of 'separable' vs 'asymmetric separable' processings is not differentiated in practice in terms of the underlying metrics, there is no way of knowing which of the two processings is adopted by the subject by way of examining the stress values.

However, in the present research we will reserve the hypothesis of asymmetric separable processing. This is because the dominant processing is obviously distinct from the separable processing (see below). Assuming that the subject happens to adopt an asymmetric separable processing due to the necessity of speedy processing, he is only attending to one compelling dimension, but if the subject chooses to attend to all relevant dimensions, we know that
the subject is to use a separable processing. That is, we can predict such a prototypical situation in which the subject's processing is not biased by a performance factor. On the other hand, we cannot make such a prediction yet in the case of the dominant processing (see §9.3.2). As we have seen in chap. 4, the separable processing requires separate attention to all relevant dimensions. It has also been characterised by a mental summation of dimensions, as is illustrated by Torgerson in the following quotation. Torgerson's hypothetical example below deals with stimuli whose structural constituents are of the three dimensions of brightness, size and shape:

It looks "as if the difference of a stimulus pair were a straight sum of the differences on the separate dimensions; that is, as though the subjects were saying these two stimuli differ this much with respect to brightness, plus this much with respect to size, plus this much with respect to shape" (Torgerson, 1958:254).

When the subject is engaged in an 'asymmetric separable' processing, it is not apparent to the experimenter whether the subject actually is capable of achieving this mental summation of components, because the subject is merely attending to one of the components currently. But the term of 'asymmetric separable' requires that it is a member of the class 'separable'. It suggests that in response to the prototypical situation which does not involve any performance factors (such as speedy processing and non-exhaustive search), the subject would process all dimensions separably to sum them up -- Goldman & Homa (1977) describes this mental summation as counting features. When we say a prototypical situation, we refer to the situation in which a given task demands the subject to attend to all relevant dimensions which are contained in his competence (see chap. 3 and below). The subject's current adoption of asymmetric separable processing implies in theory that he is potentially capable of adopting a separable processing in response to the prototypical situation. As we have seen in chap. 3, the prototypical
situation. As we have seen in chap. 3, the prototypical situation elicits all relevant knowledge of language and what is stored as accessible in one's memory are all available to him.

We have indicated three points in relation to the asymmetric separable and dominant processings. (1) There is a strong affinity between them in terms of our conceptual characterizations. (2) On the other hand, there is not any explicit definition of 'asymmetric separable' processing in terms of its underlying metrics which can usefully differentiate 'separable' from 'asymmetric separable'. (3) The underlying metrics for both 'asymmetric separable' and 'separable' processing being identical, the examination of the stress value cannot determine which of the two is adopted by the subject. However, we have also seen that the prototype for asymmetric separable processing being separable, the latter is well-defined. Because of the affinity mentioned in point (1), we may eliminate the asymmetric separable processing and might as well regard it as a special case of the dominant processing. But since the separable processing to which the asymmetric separable processing belongs is explicitly defined, it is not appropriate to eliminate the hypothesis of 'asymmetric separable' type. For this reason, we retain the hypothesis in this study. We will see in §9.3.2 that there is a good reason in regarding the asymmetric separable as distinct from the dominant processing.

Arnold (1971) and Shepard (1974) have proposed that unit circles (i.e., isosimilarity contours) can identify underlying metrics directly in the cases of the supremum and city-block metrics (see also Arabie & Boorman, 1973). The following citations explain Arnold's view, and Shepard states the same (1974:402-3 & 408):

"The argument may be clearer from this point on if the concept of the "isosimilarity contour" (a term used effectively by Shepard, 1960, 1964a, and elsewhere) is introduced. In the two-dimensional city block
model, the contour that marks all the points equidistant from a given point is a diamond shape with equal length axes parallel with the horizontal and vertical reference axes of the space. The isosimilarity contour in two-dimensional maximum component (i.e., dominant processing in the present research) space is a square."


"It may be observed in dissimilarity of meaning results obtained assuming Euclidian distance in a way that suggests the dissimilarities ought to have been scaled using either maximum component or city block distance. When Euclidian results are graphically plotted, the configurations usually conform to slightly distorted diamond shapes or parallelograms. These are, of course, the shapes of isosimilarity contours in city block and maximum component spaces."


The first citation states that the unit circle for the city-block metrics is diamond-shaped, and that the unit circle for the supremum is a square. The second citation refers to the deliberate use of the Euclidean metrics in cases where the underlying metrics of a given set of data would not be Euclidean. The shape of unit circles for the city-block metrics under this false assumption of the Euclidean metrics would be "slightly distorted diamond shapes", and the corresponding unit circle for the supremum would be "parallelograms". We will use this idea of unit circles for the analysis of our data. If we happen to obtain configurations which resemble these unit circles, we can identify which metrics the data represents. In this way, we can determine whether the subject has adopted a separable, asymmetric separable or dominant processing at the time of making similarity judgements.

The above procedure (i.e., the use of the unit circles) is chancy for the following reason. From the viewpoint
of the user of MDS, the unit circle suggests that all stimulus points converge at the 4 points (i.e., four corners of diamond or square.) The representation in which stimulus points are highly converged has been called "degenerate". In other words, the above procedure is effective only when the MDS result happens to be highly "degenerated". For this reason, the above procedure is chancy. However, it is worthwhile to adopt the notion of unit circles, since as we have noted above, in terms of our notional understanding of asymmetric separable is highly similar to the dominant processing. It is, thus, important for us to be able to differentiate these processings, particularly in our situation, since we cannot achieve sufficiently high value for r under the current restriction of the MINISSA programme, the maximum for r is 12.0 which is a far cry from \( \infty \).

Apart from the configural processing, we have dealt with a class of stimuli whose structures can be adequately represented by MDS. As Shepard (1974) makes clear, this means that a given set of stimuli involves some continuous physical variations (therefore, eligible to the quantitative dimensions in the MDS representation) or a given set involves some continuous quality which may be legitimately regarded as translatable into some numerical quantity. The former case refers to perceptual and physical stimuli such as colour, sounds, visual forms. The latter case refers to Kellerman's concrete/abstract continuity or to basic/nonbasic continuity in the present study. It may be said that when linguistic materials are used, the experimental situation typically falls into this latter case. Shepard indicates that MDS may not be effectively employed for linguistic materials, because the stimuli of "a more conceptual, linguistic or semantic nature appear to be inherently more discrete, categorical or bipolar" (1974:411-2). It must be noted that the conceptual features which are discrete or categorical are not in theory eligible in the MDS assumption of continuous quantity. For this reason, the linguistic stimuli which
typically involve discrete or categorical features pose difficult problems for MDS. For instance, a feature (+ animate) is largely a matter of whether a given entity is animate or not. There does not seem to anything in between or continuous.

Although Shepard does not give any specific name to this structure which involves categorical or discrete features, we will call this "categorical structure" here. Following Garner (1974a, 1974b, etc.) we will assume that there is a specific form of processing corresponding to a class of structure, and will call this processing 'categorical' here. Although Shepard does not define any underlying metrics for the categorical structure, we will tentatively suggest relevant characteristics for our present purposes. The following discussion is based upon Shepard (1974) and our preliminaries discussed in chap. 3.

With respect to the representational technique for categorical structure, Shepard regards a hierarchical cluster analysis modified by cross-classificatory considerations as satisfactory. For this purpose, Cunningham and Shepard have developed a technique called "maximum variance non-dimensional scaling" in which the subjective distances are measured by the "addition" of distinctive or criterial properties (1974:413). Shepard characterises the maximum variance non-dimensional scaling as follows:

\[
S_{ij} = \sum_{k=1}^{n} w_k p_{ik} p_{jk},
\]

where \( p_{ik} = \begin{cases} 1 & \text{if object } i \text{ has property } k \\ 0 & \text{otherwise} \end{cases} \)

where \( w_k \) is a non-negative weight representing the psychological salience of property \( k \).


Shepard also writes as follows:

"the perceived similarity between any two objects is a simple sum of the psychological weights"
associated with all and only those (discrete) properties that the two objects both share". (Underlying added.) (idem.)

First of all, it is important to recall our earlier discussion in chap. 3. In chap. 3 we recognized the importance of two concepts (i.e., the minimal between-category similarity and the maximal within-category similarity) in clarifying the process of acquiring word-meaning. We have tentatively defined the concept of "minimal between-category similarity" in terms of the least upper bound (lub) in a lattice theory. Similarly, we have defined "the maximal within-category similarity" in terms of greatest lower bound (glb) in a theory of lattices. These definitions are for the purpose of giving explicit definitions to experimentally and empirically well-attested notions. But also, the definitions coming from a theory of lattices is convenient for another reason. That is, Piaget's "accommodation" and "assimilation" have been widely accepted as an explanation for human learning. Since Piaget also uses a theory of lattices, our definitions are convenient in this sense as well.

The idea of lub is explained in chap. 3 as the sum set of semantic features, since mathematicians explain this lub frequently by using a simple set theory. Now we notice that our operational definition of the "minimal between-category similarity" is in keeping with Shepard's equation (3). As the second citation above reveals, in simple cases involving finite numbers of semantic features, the equation (3) stands for the sum of the shared features between any pair of objects, that is, the sum set of these shared properties. As we have already seen in reference to the truth-table in §3.4, the notion of sum set does not mean that all features are verified in a given pair of objects. It is more like Tversky's definition of metaphorical use of language. Eq. (3) above, as in Tversky's definition (see §3.4), represents the cases in which some features are shared by two objects, but it is not necessary for all features to be shared by two objects. Eq.(3) can deal with an infinite number of features (i.e.,
m=\nabla), but for our purpose we can ignore such special cases involving potentially infinite features. It is customary, however, for a mathematical expression to accommodate potentially possible cases involving an infinite number of things.

We have recalled in the above that Rosch's concept of "maximal within-category similarity" in terms of the greatest lower bound (glb). In chap. 3, glb is explained in reference to set intersection (∩ or \Pi). Using Shepard's equation (3), if we suppose that all properties in question are shared by two objects, we get

\begin{equation}
S_{ij} = \prod_{k=1}^{n} w_{ik} p_{ik} p_{jk}.
\end{equation}

Eq. (4) reveals that our operational definition of "the maximal within-category similarity" can be accommodated by the equation derived from Shepard's original equation.

It is important to notice that the two notions of "maximal between-category similarity" and "minimal between-category similarity" are central to Rosch's prototype model of word-meaning. More important, our operational definitions of these concepts are in keeping with Shepard's definition of categorical structure. For this reason, we may regard categorical structure as the likeliest candidate for representing word-meaning among the five kinds of structures we have considered here.

However, we have already noted that Shepard does not offer any underlying metrics for the categorical structure. In the context of the present discussion, one can guess more readily why Shepard has not specify any metrics for this structure. Prototypes themselves (p_{ik} p_{jk} in eq. 3 & 4) are difficult to define, since the definition depends upon individual cases. In some words, their main functions may be to enable the speaker and the hearer to identify an object named, such as an apple. But the meaning of an apple involves not only physically specifiable properties of shape and colour, but also nonphysical
The difficulty in specifying metrics is due to the fact that words involve this mixture of physical and nonphysical properties. Even when we restrict our discussion to physical properties, we notice that each word tends to refer to different shapes and colours. For the perception of colour, the Euclidean metrics has been regarded as suitable (Indow & Shiose, 1956 & 1958, Indow & Kanazawa, 1960, Indow & Matsushima, 1969, Indow & Ohsumi, 1972, etc.) Whereas the simple shape such as schematic faces (which are far simpler than complex forms in our daily life) tend to be associated with the city-block metrics; Goldman & Homa (1977), Rosch (1976), etc. Since words tend to involve the more complex shapes and colours, we will find it difficult to determine metrics even in the case of physical properties. The situation is worse, when we consider more abstract properties in conjunction with physical properties. Arnold's experiments (1971) suggest that the words Arnold dealt with conform, by and large, to the supremum metrics rather than Euclid, but that the different parts of speech require somewhat different metrics. This casual example seems sufficient to illustrate the difficulties in specifying underlying metrics or norms for word-concepts.

Eq. (3) and eq. (4) appear to be relatable to Piaget's view of cognitive development from "modular" to "complemented and distributive" lattices. (The examination of the relationship clarifies the nature of categorical structure in relation to the definition of word-meaning. The discussion will lead to an important characteristic of categorical processing by which we can identify the categorical processing. ) By the time Piaget's INRC (§3.4) is attained by the subject, his cognitive operation is highly similar to a "complete" Boolean algebra (Piaget, 1953, Flavell, 1963, Hatano, 1965, etc.). This thesis does not offer any detailed comparison between Piaget's use of logical apparatus and Boolean algebras. We will merely point out two important basics which are relevant to the
distinction between competence and performance in the area of word-meaning. These two basics are traditionally called "duality principle" and "complete" in mathematics. The purpose of this discussion is to appreciate the relationship between eq. (3) and (4). This relationship not only characterises the categorical structure adopted here, but is also relevant to the distinction between competence and performance.

Piaget does not use such symbols as $\Sigma$ and $\Pi$ which are used in eq. (3) and (4). As we have already noted, a mathematical expression tends to accommodate some hypothetical infinitude; for our present purpose it does not make any difference whether eq. (3) & (4) are rewritten as below, adopting Piaget's notations. We will ignore $w_k$ which is a positive weight representing the psychological salience of a given property $k$ (see eq. (3)).

\[
(3') S_{ij} = \sum_k P_{ik} P_{jk} \text{ or simply } P_{i \wedge k} + P_{j \wedge k}.
\]
\[
(4') S_{ij} = \prod_k P_{ik} P_{jk} \text{ or simply } P_{i \vee k} \times P_{j \vee k} \text{ or } P_{i \wedge k} ' P_{j \wedge k}.
\]

Eq. (3') & (4') show that the basic operations in eq. (3) and (4) correspond to the Boolean operations of $\wedge$ and $\vee$. Piaget regards cognitive development as progressing from modular to complemented and distributive lattices. This means that the cognitively less mature subjects show a marked tendency to rely on eq. (3') and (4'). Whereas, the cognitively mature subjects show in their formal operations both eq. (3') and (4') hold simultaneously - this is due to the "duality principle" (ibid). As we see below, this characteristic of the latter has some relevance to our characterization of the relationship between competence and performance (see below).

Eq. (3') and (4') represent the respective relationships between operators $\Sigma$ and $\wedge$, and between $\Pi$ and $\times$ (or $\cdot$), according to the standard conventions in theories of lattices and Boolean algebras (e.g., Takeuti, 1971 and Matsumoto, 1977). In the above, all the equations are based upon the similarity between two objects 'i' & 'j'. However, the representations in Boolean algebras and
So far we have adopted the "intensional" definition of word-meaning, as opposed to the "extensional" definition. That is, \( p_{ik} \) or \( p_{jk} \) refer to the semantic properties the two objects share. This "intensional" definition is distinct from the extensional definition in which the elements belonging to a given word-concept are listed. In order to show that a Boolean algebra is useful in specifying a single word concept, it is convenient to adopt the extensional definition of word-concept at this point. The difference of extensional or intensional definition does not matter in the present case, because in the following explanation of the "complete" Boolean algebra in which the greatest or least element exist, the corresponding "intensional properties" have been proved to exist via Zorn's lemma (see Takeuti, 1971:218-22; Seki, 1959, appendix). So, once the greatest and least elements are specified, the definite intensional properties which determine the membership of a word-concept are in theory in existence.

Takeuti (1971:89) explains the nature of "complete" Boolean algebra as below. We notice below that the two notations (\( \Sigma \) and \( \Pi \)) are convenient in specifying the presence of the least as well as greatest elements in any subset \( \{b_i \mid i \in I\} \) which belongs to a set \( B \).

"Definition: Let \( \{b_i \mid i \in I\} \) be a subset of \( B \). If there is the least element in \( b \), we write this least element as

(5) \[ \sum_{i \in I} b_i \quad \text{where} \quad \forall i \in I \quad (b \geq b_i) \]

Also, if there is the greatest element in \( b \), we write this greatest element in \( b \) as

(6) \[ \prod_{i \in I} b_i \quad \text{where} \quad \forall i \in I \quad (b_i \geq b) \]

If any subset of \( B \) (\( \{b_i \mid i \in I\} \)), if there is both \( \Sigma b_i \) and \( \Pi b_i \), \( B \) is said to be complete." (\( B \) stands for \( \langle B, +, -, 0, 1 \rangle \)). Takeuti, (1971:88.)
As the above quotation makes clear, the use of operators \((\Sigma \, \& \, \Pi)\) makes Piaget's expressions more general. We can also notice that the above representations \((\Sigma_{\leq k} \, \& \, \Pi_{\leq l} b)\) specify the least element and the greatest element respectively. Since the former is the least element in the upper bound \((b \geq b)\) and the latter is the greatest element in the lower bound \((b \leq b)\), each stands for the least upper bound \((lub)\) and the greatest lower bound \((glb)\) respectively.

According to the above standard definition of complete Boolean algebra, the presence of lub and glb in fact determines the complete nature of a given set of objects. We recall that these lub and glb give the present definition of word-concept, the former defining the minimal between-category similarity, and the latter, the maximal within-category similarity. To relate the above discussion to this suggests that the present definition of word-meaning presumes complete lattices or Boolean algebras.

It has been known that when the number of elements belonging to a set is finite, this set of objects necessarily forms a complete lattice or Boolean algebra (ibid.) There are in theory innumberable objects, events or states which a specific noun, verb or adjective can refer to. However, we recall that there are always superordinate concepts which categorize these specific nouns, verbs and adjectives. Thus, we realize that each word refers to a limited number of things (events and states), when we have superordinate concepts in mind. In this sense, we can say that a word denotes a finite set of abstract or concrete entities. If this is the case, it suggests that our assumption of complete lattices or Boolean algebras is justified.

When it is reasonable to assume that the definition of word-meaning conforms logically to the complete nature, the above definition can be applied to a single word rather than a pair of words (see above). This is because when a given lattice or Boolean algebra is complete, the lub is equivalent to the maximal element and the glb, to the
According to our previous discussion (chap. 3 and §9.2), the word-meaning of break_2 is determined by the two word-boundaries \( \xi_1 \) & \( \xi_1 \). This previous definition involves the other two words, 'break_1' & 'break_3'. However, in the case of a complete lattice or Boolean algebra, the lub is equal to the least element which is properly included in break_2. At the same time, the glb which belongs to break_1 is equal to the greatest element which is a proper element of break_2. For this reason, the definition through lub and glb in fact offers the definition of break_2 itself. This example makes it clear that the present definition of word-meaning applies to both a pair of words and a single word on its own.

The "duality principle" characterises the relationship between eq. 5 & 6 (hence, eq. 3 & 4), as below. Let us take the most relevant example of the "duality principle" from Takeuti (1971:74-5):

\[
\begin{align*}
(7) & \quad b_1 + (b_2 + b_3) = (b_1 + b_2) + b_3 \\
(8) & \quad b_1 \cdot (b_2 \cdot b_3) = (b_1 \cdot b_2) \cdot b_3
\end{align*}
\]

In this example, the duality principle indicates that when (7) holds, (8) holds simultaneously, and vice versa. At this point it is relevant to recall that eq. (7) is a special expression of the general expression, eq. (5). Likewise, eq. (8) is an instance of the more general expression, eq. (6). Further, in eq. (3) & (4), once we keep \( w \) constant (i.e., we get eq. (3') & (4')), we have already noted that eq. (3) & (4) correspond to eq. (5) & (6) respectively. Since the duality principle exists in eq. (7) & (8), the same duality principle exists in other paired expressions.
Then, because of the duality principle, we can say that eq. (3) is triggered in cognition, eq. (4) would simultaneously hold in cognition and *vice versa*.

When we translate the above "duality principle" in the context of the present research, we can infer two points. The first point relates to our adoption of Rosch's concept of word-meaning. We have argued that eq. (3) & (4) represent the present definition of Rosch's "minimal between-category similarity" and "maximal within-category similarity" respectively. The "duality principle" suggests that once the subject estimates either of these criteria, he automatically attains an estimation of the other. This may suggest to us that as far as we are concerned with the subject's performance in using words, the availability of one criterion ( lub or glb) is sufficient. If only one of these two is verified in the subject’s immediate context of speech, the two necessary lexical criteria are naturally satisfied according to this "duality principle". The same argument will hold not only for pairs of words but also for single words, since the same set of expressions, eq. (5) & (6) is applicable to the definition of a single word, as we have seen above.

At this point it is useful to remember our characterization of competence and performance (see chap. 3). The distinction between them is made in chap. 3 on the basis of one simple point, that is, the availability of all or some semantic features in the immediate context of speech. We have argued that competence refers to the situation in which all relevant semantic features are available and verifiable. This is the reason why the glb is offered as the likely logical expression for representing this sense of competence. Whereas, when some features are available in the immediate context of speech, the context triggers performance on the part of the subject. Thus, performance appears to be best expressed as the lub. In this section we have seen that the definition of the glb and lub correspond to Shepard's view of categorical structure. We have also ascertained that in defining categorical struc-
ture this way, the complete lattices or Boolean algebra can account for the categorical structure seen in the area of word-meaning. Now, according to the "duality principle", the verification of either the lub or the glb leads automatically to the verification of its counterpart. The available lub makes the lub accessible and vice versa. This relationship is highly significant in the competence vs performance distinction. This is, competence is assessible through performance, since when the lub is satisfied, the glb is satisfied accordingly.

Additionally, the simple criterion of all or some features may bring about the following implication in the present discussion of the five kinds of processings: (1) integral; (2) separable; (3) asymmetric separable; (4) dominant; (5) categorical. It is clear that the first four processings would not require all relevant semantic features. If this is the case, these four kinds of processing are more intimately related to performance rather than revealing competence (see §9.2.3 (2)).

The proper justification of the adoption of lattices and Boolean algebras is beyond the scope of this thesis. We will gain some credulity in noting one point. The youngest subjects who participated in the present experiment are 14 to 15 years old. According to Piaget's developmental schema, this means that they have attained the "formal operational period". This means that the youngest group has reached the stage of manipulating "distributive complemented" lattices or the complete Boolean algebra. This supports our application of these branches of logic here.

On interpreting \( \Sigma \) and \( \Pi \) as denoting the operations in Boolean algebra, it is called a topological Boolean algebra. In this topological Boolean algebra, it is no longer necessary for us to restrict our discussion to the area of Shepard's "non-dimensional scaling". Since \( \Sigma_{b_i} \) and \( \Pi_{i} \) belong properly to topological spaces (Takeuti, 1971), we can thus connect our discussion with the issue of metrics in multi-dimensional spaces. In this sense, it is
appropriate for us to use MDS even in the case of categorical structure, if we are interested in looking at metrics (norms) which identifies the categorical processing.

When the equivalence is defined as below, some mathematicians have proved that various metrics \( r = 1, 2, \ldots, \infty \) are equivalent, particularly when the structure in question satisfies the requirements of distributive lattices (e.g., the complete Boolean algebra); Hirsch & Smale, (1974: chap. 5, § 2 propositions 1 & 2); Kobayashi, (1977: chap. 5); Takeuti, (1971:101, theorem 16). According to Kobayashi (1977:107-8), the equivalence of distances (norms or metrics) is defined as follows:

"Theorem 3: the necessary and sufficient conditions for the two distances, \( d \) & \( d' \), to be equivalent in a space \( X (\neq 0) \) are as follows:

(1) There is a neighbourhood \( U(x; \delta; d') \) such that for any \( x \in X \) and for any \( \epsilon \) in \( U(x; \delta; d) \) (\( \epsilon > 0 \)),

\[ U(x; \delta; d') \subseteq U(x; \epsilon; d) \]

(2) There is a neighbourhood \( U(x; \delta; d) \) such that for any \( x \in X \) and for any \( \epsilon \) in \( U(x; \epsilon; d') \) (\( \epsilon > 0 \)),

\[ U(x; \delta; d) \subseteq U(x; \epsilon; d'). " \]


Hirsh and Smale mention that "every norm for any subspace on \( E \) is obtained from a norm on \( R^n \) by restriction where \( E \subset R^n \)": Hirsch & Smale, (1974:79). Likewise, given a norm, any subspace representing a given categorical structure can be obtained by restriction.

Hirsh and Smale's and Kobayashi's proof has one important consequence for us and we can make two additional comments which are relevant to the present discussion. (1) Hirsh and Smale, and Kobayashi have proved that the various metrics (distance in Kobayashi's terminology and norm in Hirsh & Smale) are equivalent. We notice that according to Piaget's developmental scheme the cognitively mature
subjects exhibit the formal operations of distributive lattices. As we see below, the categorical structure can be more properly defined by the simultaneous operation of eq. (3) and (4). Since the norms are proved to be equivalent, whatever parameter we choose, the stress values yield no significant differences, if the subjects adopt this categorical processing. This prediction appears to be adequate in that the stress values are the measure of discrepancy between input and output (see chap. 5); when the various metrics serve as equivalent, and the input remains the same, the measure of discrepancy remains equivalent to whatever metrics we may choose. That is, by adjusting parameters variously, if the stress values turn out to be equivalent we can conclude that the subject has adopted the categorical processing at the time of his making his similarity judgement. We will have to solve the issue of what the equivalence is in practice, since the notion of equivalence above is a logical one. We need to translate this logical notion into some practical means including some rounding errors in §9.3.2.

Two additional comments can be made in relation to the above proof. Shepard (1974) emphasises that the Euclidean metrics is "robust" in representing whatever structure a given object happens to have. The proof mentioned above certainly appears to support Shepard's view in that when metrics are logically equivalent, the Euclidean metrics would serve for all purposes (see Diagrams 11 & 12, in chap. 10). Secondly, with respect to the definition of equivalence, the MDS methodology is justified in reducing the number of dimensions, as long as a given structure preserves the equivalence, for instance, between the \( U \) in a 4-dimensional space and \( U \) in a 3-dimensional space.

In this section we have overviewed the 6 kinds of processing following closely Garner's and Shepard's characterizations. It is particularly important to bear in mind the following two points in the present experiments presented in this chapter. (1) The four processings (integral, separable, asymmetric separable and dominant) have been characterised
by the specific underlying metrics. In the case of the categorical processing, it has been argued above that whatever parameters we supply as underlying metrics, the results in terms of stress values will be equivalent. (2) We have also seen that the unit circles can determine a form of processings the subject has adopted at the time of his making similarity judgements in some cases. The following table summarizes the present discussion.

<table>
<thead>
<tr>
<th>metrics</th>
<th>unit circles</th>
</tr>
</thead>
<tbody>
<tr>
<td>integral</td>
<td>Euclid</td>
</tr>
<tr>
<td>separable</td>
<td>city-block</td>
</tr>
<tr>
<td>asymmetric</td>
<td>city-block</td>
</tr>
<tr>
<td>dominant</td>
<td>supremum</td>
</tr>
<tr>
<td>categorical</td>
<td>equivalent</td>
</tr>
</tbody>
</table>

Table 2 Processings & Metrics

§ 9.3 Experiment

This section presents the method of experiments in § 9.3.1, and the experimental results in § 9.3.2.

§ 9.3.1 Method

As we have seen in § 9.2, we follow two procedures.

(1) Since the specific form of processings relates closely to the three kinds of metrics (with the exception of the categorical processing), we will examine which metric can yield the best goodness-of-fit to the input data, i.e., the lowest 'stress' values.

(2) If we happen to obtain what is normally regarded as highly "degenerate" data, we would examine whether the unit circles corresponding to city-block and supremum metrics can be observed in the output configurations from MDS. Additionally, we have noted the possible implications of Hirsh and Smale's and Kobayshi's proof about the equivalence of norms in relation to the categorical structure. In the case of the categorical processing, what-
ever parameter we supply as a possible underlying metric, the results in terms of the stress values would be equivalent. We have tentatively argued that this equivalence can prove the possible operation of the categorical processing.

As we have noted previously, the first procedure requires us to balance the problem of 'dimensionality' against the issue of changing parameters. That is, we look for the lowest stress value by manipulating both 'the number of dimensions' and 'parameters'. This is a tricky procedure. For this reason, it is more convenient if we can predetermine one of the two factors prior to the experiments. There has been suggested a guideline for the problem of dimensionality. Edinburgh regional computer centre writes as follows:

"the ratio between the number of data elements and the number of latent parameters* should be at least two." (Edinburgh Regional Computer Centre, 1981: 8-9.)

*Parameters here stand for coordinate axes.

Since there are 55 or 45 data elements in the present experiments, the number of coordinates in the 2-dimensional space is 22 or 20. Since the compression ratio should be at least 2, the appropriate number of dimensions here is 2 or 3. Just in case we will run MINISSA program on the basis of 2-5 dimensions. We will examine below which underlying metric would yield the lowest stress values. We will choose r=1.0 (cityblock), 2 (Euclid), ..., 12.0 (which is the maximum value we can supply under the current restriction of Fortran). The last point is indicated by J. Muxworthy, in ERCC.

As we have noted in §9.2, the second procedure of the unit circle is chancy. Since it might be possible to obtain results that resemble the unit circles (i.e., diamond in the city-block metrics, and square in the supremum metrics), this procedure is adopted here. According to this procedure,
we only need to obtain a 2-dimensional solution. We also see whether the input data can conform to a slightly distorted diamond or parallelogram, under the false assumption of the Euclidean metrics, adopting Arnold (1971) cited earlier.

As for the categorical processing, we have tentatively argued that whatever metric we supply as a possible candidate for an underlying metric, the stress values would be equivalent. We will see whether this tentative argument is real with the data we have obtained here.

The same five groups of subjects who took part in the experiment of translatability judgements (chap. 10) participated in the present experiments. They made the judgements of similarity among possible combinations of paired sentences, following the elicitation method of the categorical sorting explained in chap. 5. The scale presented to the subjects is a 6-point scale, the use of which is operationally justified in chap. 5. The ratings are tabulated according to each pair. Then, adopting the Thurstone-Torgerson method which is reinterpreted in chap. 7, the ratings are converted into SVs. These derived SVs are then ranked. The resulting rankings are the input data for MINISSA. The reason why these derived SVs are used to obtain the final ranks has been presented in chap. 5, where we have detailed some advantages over the other methods such as card-sorting or straightforward paired comparison.

Since the subjects are grouped according to their school years, we may examine whether the groups of subjects adopt different types of processings according to their ages. Although the school years may not reflect their cognitive development directly, the results may have some developmental implications.
§ 9.3.2 Results & discussion

As we have seen in chap. 5, "stress1" estimates whether the MDS results are fair or poor in relation to the ordinal information supplied as input. Stress1 is computed, assuming random data and depending on the number of data points, stress1 varies accordingly (Spence, 1979). In the present experiments, one group (3rd year junior high school students dealing with the yaburu sentences - coded as YAC3 below) is exposed to the 10 sentences. On the other hand, the other groups deal with the 11 sentences. Thus, YAC3 yields a different stress1 from the other groups:

<table>
<thead>
<tr>
<th>YAC3</th>
<th>The Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim</td>
<td>dim</td>
</tr>
<tr>
<td>5</td>
<td>0.122327</td>
</tr>
<tr>
<td>4</td>
<td>0.116557</td>
</tr>
<tr>
<td>3</td>
<td>0.122474</td>
</tr>
<tr>
<td>2</td>
<td>0.154917</td>
</tr>
</tbody>
</table>

Table 3 Stress1

The 10 sets of results obtained yield considerably lower stress values than the stress1 above (see Appendices 6.2.1 tables for stress values. In terms of the stress1 the present results are guaranteed not random.

The specific results in relation to the underlying metrics are presented below. (1) we will see which metric can account for the learner-oriented input data; i.e., we will examine which metric yields the lowest stress value of all. This examination involves some practical issues which are not touched on in §9.3.1. This is the problem of tradeoff between dimensionality and metrics, which is briefly mentioned in general terms in §9.2. The practical procedure adopted is presented here. (2) In the case of the categorical processing, we have seen that no specific metric could not be attributed to. In reference to Hirsh & Smale's and Kobayashi's proofs, we have tentatively stated that the logical notion of equivalence holds among various
metrics. We still have the issue of how we can translate this logical notion of equivalence into some arithmetic value. This issue is also discussed in practical terms. The results are presented below on the basis of the present operational procedures. (3) The two groups (YAC3 & KOU2 ie., 2nd year university students involving the kowasu sentences) yield highly 'degenerate' results which resemble the unit circles of the Euclidean and city block metrics respectively. This result is discussed in (3) below. (4) The developmental pattern in terms of processing and the MDS representation of the NL states is also discussed as the fourth topic in this section.

§ 9.3.2 (1) dimensionality and metrics

MINISSA computes two kinds of stress: \( \hat{d} \) & \( d^* \). \( \hat{d} \) is relevant when the procedure called "hard squeeze" is employed during the computational process. Whereas \( d^* \) is relevant when the "soft squeeze" is consistently adopted in the process. As we have seen in chap. 5, MDS attempts to 'squeeze' data points in as low a dimensional space as possible; in MINISSA, it employs a 'soft squeeze' first and later a 'hard squeeze', only when the resulting configurations fit poorly to the input rankings.

Table in 6.2.2, Appendices lists which set of data has required a soft or hard squeeze in the computational process. According to this result, the relevant stresses in each group of subjects are ticked in tables, 6.2.1, Appendices.

There has not been suggested any 'rigorous' rules for determining how adequate a given stress value is. However, since Kruskal (1964), there has been accepted a 'rule of thumb'. We adopt a guideline proposed by inter-university /research councils (1977) which is almost the same in Kruskal (1964:3).

\[ \text{stress} < .01 \quad \text{excellent} \]
According to the above proposal, we will assume that the notional criteria of 'good' or 'excellent' is adequate for our present purpose. This means that a given stress value must be lower than 0.05.

The subjects deal with 10 or 11 sentences. As we have seen in §9.3.1, 2 or 3 dimensional spaces are mathematically appropriate in representing these 10 or 11 data elements. On the basis of the above rule of thumb concerning stress, one can check again whether the assumption of 2 or 3 dimensional spaces is adequate in terms of stress. This extra check is desirable, since contrary to the mathematical restriction the subjects may have in practice relied on more than 3 continuous semantic features which are indeed in the MDS represented as dimensions. Although 6.2.1 in Appendices lists all the necessary stress values, we look at a relevant part here.

<table>
<thead>
<tr>
<th>2 dimensions</th>
<th>3 dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>r = 1.0</td>
<td>r = 1.0</td>
</tr>
<tr>
<td>r = 2.0</td>
<td>r = 2.0</td>
</tr>
<tr>
<td>supremum</td>
<td>supremum</td>
</tr>
<tr>
<td>KOC3 0.085978</td>
<td>0.085978</td>
</tr>
<tr>
<td>KOK1 0.072268</td>
<td>0.072268</td>
</tr>
<tr>
<td>KOK2 0.085862</td>
<td>0.085862</td>
</tr>
<tr>
<td>KOU1 0.037370</td>
<td>0.037370</td>
</tr>
<tr>
<td>KOU2 0.000005</td>
<td>0.000005</td>
</tr>
<tr>
<td>YAC3 0.000005</td>
<td>0.000005</td>
</tr>
<tr>
<td>YAK1 0.000005</td>
<td>0.000005</td>
</tr>
</tbody>
</table>
The sets of data which satisfy the present criteria ($s < 0.05$) are ticked in Table 5 above. We obtain the following results, when the dimensionality is determined by the present method.

<table>
<thead>
<tr>
<th></th>
<th>KOWASU</th>
<th>YABURU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KOC3  KOK1  KOK2</td>
<td>YAC3  YAK1  YAK2</td>
</tr>
<tr>
<td></td>
<td>3     3      3</td>
<td>2/3 2/3 2/3</td>
</tr>
</tbody>
</table>

Table 5 Stress

Table 6 Dimensionality

Thus, one can specify according to the practical rule of thumb the appropriate number of dimensions. The next step to follow naturally is to see which metric has yielded the lowest stress as regards to each of the ten sets of data. Under usual circumstances, this procedure might have been effective. We however notice at once that, of those ticked as satisfactory in Table 5, the differences in the stress values are almost negligible. The differences are all far lower than 0.05. Furthermore, the stress values yielded on the assumption of the city-block metric are precisely the same as those yielded on the assumption of the Euclidean metric. This salient feature is consistent across the five groups of subjects and in the two verbs dealt with here. Irrespective of the number of dimensions, this salient feature is observable (see Tables in 6.2.1, Appendices). Unless we take some consideration of this feature into account, the underlying metric may not be fully determined. For this reason, we will postpone the specification of metrics on the basis of the lowest stress under the above dimensionality.
Table 6), until we look into the categorical processing in (2) below.

§ 9.3.2 (2) specifying a form of processing based on stress

In §9.2 we have argued on the basis of Hirsh and Smale's and Kobayashi's proofs that the categorical processing would yield "equivalent" stress values, whatever parameter we choose to supply. We have chosen the three parameters: the Euclidean, city-block and supremum metrics. Table 5 and Tables in 6.2.1, Appendices indicate that this prediction appears to be born out. We have so far seen two reasons; (i) the stress values yielded by the Euclid metrics is identical to those by the city-block metrics among the ten sets of data dealt with here. The same consistency is observed in the 2 - 5 dimensional spaces. The latter point is also consistent with Kobayashi's account cited and commented in §9.2. (ii) Furthermore, it appears to look as if the differences in the stress values computed on the assumption of the three metrics were almost negligible. However, we need to see whether the logical notion of "equivalence" can be translated into something which is vaguely stated as 'almost negligible' above. Here, we will interpret the logical notion of equivalence in three ways, and see which is the most appropriate and harmonious with the present definition of word-meaning.

The three interpretations are (a) the complete identity of stress, (b) the permissible 'equivalence' involving arithmetic rounding and (c) the 'equivalence' which requires some further scrutiny into the categorical structure presented in §9.2.

(a) The **complete identity of stress** in the three metrics states that all the three metrics must yield exactly the same stress values. This condition of the complete identity is met only by one set of data, YAK1. This condition is highly restrictive, since we are considering identity to the 8th decimal place. This is particularly so, when we consider various computational procedures in MDS.
Lingoes (1975) states that in order to reproduce all the rankings in a low dimensional space, MDS goes through four types of transformations: "translation, rotation, reflection and dilation". Among others, rotation tends to alter the positions of stimulus points radically.

![Diagram of rotation](image)

Since stress is the measure of discrepancy between the initial inter-point distances and the distances derived in the lower dimensional spaces, one can understand how the process of rotation can affect stress values. Furthermore, the process of iteration can change the values in the 6th or even 5th discimal places. Table in 6.2.3, Appendices shows how many iterations each set of 10 data have gone through in the computational process. The table indicates clearly that while the Euclidean and the city-block metrics use the same number of iterations, the supremum metric employs the larger number of iterations. For these reasons, we may say that the condition of the complete identity is not a realistic translation of the logical notion of equivalence. The computational process is bound to involve a great deal of rounding errors, due to the repeated use of the set of same equations.

(b) This method called here permissible equivalence allowing for rounding errors is self-explanatory. The rough rule of thumb such as seen in Table 4 may be more realistic in translating the logical sense of "equivalence" into approximations. Since the stress values listed in Table 4 has been accepted well over two decades, we will also regard approximations which occur at the 2nd decimal place as fair here. According to this definition, the following
7 sets of data are regarded as satisfying the condition of equivalence, and hence they are regarded as conforming to the categorical processing.

<table>
<thead>
<tr>
<th>KOC3</th>
<th>KOK2</th>
<th>KOU1</th>
<th>KOU2</th>
<th>YAC3</th>
<th>YAK1</th>
<th>YAU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2D)</td>
<td>(3D)</td>
<td>(2D)</td>
<td>(2/3D)</td>
<td>(2/3D)</td>
<td>(2D)</td>
<td>(3D)</td>
</tr>
</tbody>
</table>

Table 7

This criteria of equivalence cannot account for why the Euclidean and city-block metric yield identical stress values consistently. This consideration leads to the third interpretation of equivalence (c) below.

(c) The categorical structure (and hence the categorical processing itself) may be given some modification to explain the complete identity of stress observed in the Euclidean and city-block metric. First of all, this consistent result suggests that all the subjects participated in judging a given set of experimental items integrally and separably. Integral processing is a wholistic grasp of 'semantic features' in the present situation. Whereas, Torgerson's account cited earlier shows that separable processing involves the counting of such features. In view of the categorical structure defined in §9.2, we notice that we can relate the wholistic grasp and counting to the respective operations of \( \Sigma \) & \( \Pi \) in eq. 3 & 4 or 5 & 6. It may not be the only way, but a global grasp is made possible by drawing the subject's attention to the intersection of features, ie., \( \Pi b \). Also, \( \Sigma b \) is plainly stated as a summing (i.e., counting) of features. As far as the basic operations in lattices or Boolean algebras are concerned, \(+/\Sigma\) can represent counting which corresponds to the notion of the city-block metric and \(-/\Pi\) represents the commonality which is shared by the features. Thus, we may tentatively suggest that eq. 3 represents separable processing and eq. 4 represents integral processing. As we have seen in §9.2, eq. 3 & 4 jointly defines a word-meaning; the former defines the minimal between-category similarity and the latter, the maximal within-category similarity. Since both are
jointly necessary to define the meaning of a word, it may follow that the integral processing and separable processing must occur jointly. This seems to account for the complete identity observed in all the data examined in this study. It must be born in mind that these two processings must occur simultaneously; then it is a form of categorical processing. This further analysis of categorical structure seems to be in keeping with the operational definition of word-meaning presented here. When separable processing or integral processing occurs on its own, such data emphasizes 'performance' rather than 'competence'. But as we have seen above, if a given set of experimental items are conceived by the subjects to conform to the categorical structure, then, eq. 3 triggers eq. 4 automatically, due to the duality principle discussed in § 9.2.

According to the above interpretation, all the 10 sets of data can be regarded as subject to the categorical processing. Ideally, the three metrics ought to produce 'equivalent' stress values. The supremum metric does not behave exactly the same way as the other two metrics. However, we must recall that because of the restriction of the current MINISSA program we could not raise the value of $r$ more than $r = 12.0$. In some cases, $r = 9.0$, because the calculation causes "exponential default" (i.e., exceeding the power of Fortran). Either $r=12.0$ or $9.0$, it is a far cry from the proper sense of "supremum" ($\infty$). It is possible to think that if we attain a sufficiently high value to claim the status of supremum, the resulting stress value might have been more similar to those by the city-block or Euclidean metrics (see below).

We may tentatively conclude here that (c) can account for the identical stress values by the Euclidean and city-block metrics best of all. (c) also is consistent with the operational definition of word meaning here. Further, only (c) can explain why the stress values are identical irrespective of the number of dimensions. This phenomenon of identical stress whatever dimensionality it may take
is a natural consequence of Kobayashi's proof cited earlier. For these reasons, we adopt (c) in this study. According to (c), all the 10 sets of data conform to the categorical processing as well as structure. As we see below, categorical processing in fact requires the parallel operation of dominant processing as well.

We have seen above that the present definition of the categorical structure which is defined by the paired equations of eq. 3 & 4, or eq. 5 & 6 appears to correspond to the present interpretation of joint operation of integral and separable processings. We have ascertained that these two operations must occur jointly. The reason is briefly stated as follows. Because one of the paired expressions (eq. 3 or eq. 5) represents the minimal between-category similarity and the other pair (eq. 4 or eq. 6), the maximal within-category similarity. These two kinds of similarity in combination defines the present interpretation of word-meaning. For this reason, we are justified by saying that both integral and separable processings occur jointly in understanding the meaning of a word.

We have seen in § 9.2 that the dominant processing corresponds to the supremum metric. Should the subjects adopt a categorical processing, we would expect that the three metrics ought to yield some equivalent stress. This theoretical prediction was not born out by the present experiment, largely due to the practical unfeasability of computation under the current restriction of MINISSA. But, since Arnold (1971) manages to compute up to $r = 32.0$ and Goldman & Homa (1977), $r = 50.0$, some sort of maximum rather than supremum is indeed feasible; for this reason Arnold apparently chooses to name our dominant processing after this feature of maximum, i.e., "the maximum component" processing. However, we should not confuse the practical unfeasability with the theoretical realizability (see below). For this reason, it is necessary for us to examine whether the present interpretation of the categorical processing can include the dominant processing as its integral part. That is, whether the joint operation of
integral & separable processing serves the purpose of the
dominant processing. Specifically, we will examine below
whether operator 'Σ' in eq. 3 and eq. 5 is intimately related
to the present understanding of the dominant processing.

The dominant processing has been characterized as the sub-
ject’s reliance upon the maximum component. As we have
seen in §9.2, Arnold describes this notion as "the one
that maximally discriminates the members of a concept pair"
(1971:349). It is important to remember that the same
component must exist in all concept pairs; i.e., the same
maximum component which discriminates maximally two extremes
ought to discriminate maximally a closely related pair.
This suggests that the maximum component has the least
similarity among all concept pairs. The sense of
similarity is so little that it discriminates the members
maximally. We may easily connect this notion of the least
similarity with eq. 3 and eq. 5 each of which specifies
the minimal between-category similarity. This conceptual
intuition appears to be born out by these mathematical
definitions. The maximum component interpreted by eq.
3 or eq. 5 represents in fact some superordinate concept.
The complete nature of complemented and distributed lattices
and Boolean algebras, which is explained in §9.2, states
that the lub is equal to the minimal element belonging
to the lower category/concept (e.g., according to the
example given in Diagram 2 above, the lub in break₃ is
equal to the minimal element of break₂) and that the glb
is equal to the maximal element belonging to another lower
category/concept (e.g., the glb belonging to break₁ is
equal to the maximal element of break₂). This equality
is maintained infinitely, which is a characteristic feature
of the complete lattices or Boolean algebras. We can
ascend to the higher concept infinitely; and equally we
can descend to the lower category/concept infinitely.
Yet, we get the self-same equality. Some concept which
is shared by not only lower concepts (hyponyms) but also
by the higher category/concept is by definition a
superordinate concept, such as [+animate]. Although we
need the mediation of eq. 6 to descend to the lower
categories, eq. 5 is sufficient to represent any superordinate concept.

What has emerged in the above discussion is that the dominant processing is implicitly accommodated by the separable processing whose essense is expressed by eq. 5. To be more specific, assuming that the separable processing of the form eq. 5 is carried out infinitely, we will get a supremum. That is, the supremum is the outcome of the infinitely operated separable processing; the process in which the subject arrives at his psychological maximum component is specified by the separable processing itself. The supremum metric, we recall, specifies that a given operation is based upon the dominant processing. In this sense, the difference between the separable and dominant processings is in whether the same process operates on finite or infinite things. At the same time, since the subject depends upon eq. 6 to check whether it is really maximally discriminating the members in his mind (see above), it is most likely that the integral processing whose function is defined by eq. 6 is in operation in order for the subject to examine the truth of his unconscious cognitive activity.

Incidentally, the above discussion concerning the dominant processing and superordinate concepts brings about one consequence in our distinction between the asymmetric separable and dominant processings. As we have noted in §9.2, these processings are notionally defined similarly; i.e., the subject processes a given object, paying attention to the one compelling obvious dimension. The above discussion makes it clear that the dominant processing is best understood as closely related with superordinate concepts. Thus, it is now clear that in the case of the dominant processing, this compelling dimension stands for a superordinate concept in the mind of the subject. Whereas, in the case of the asymmetric separable processing, the one compelling dimension can be any one of several features the subject's attention is drawn to. It may happen to coincide with a superordinate feature. Because
the asymmetric separable processing is based upon the city-block metrics, this processing involves not only mental summation of features or counting, but also separate attention to individual features. In this sense, the asymmetric processing itself cannot specify what is the compelling dimension in the mind of the subject. The subject’s attention can be drawn to any feature which happens to be salient or relevant at the time of his performing a task. Thus, while the asymmetric separable processing depends on any compelling feature, the dominant processing calls for a superordinate feature. Thus, there seems to be a good reason for retaining the distinction between these processings, although we have noted the notional similarity between them in §9.2.

**Summary:**
We began our discussion by noting that the equivalence between the two kinds of stress values yielded by the integral and separable processings is sufficient evidence of the categorical processing. The discussion led to some elaboration of the present definition of word-meaning which is characterised in this chapter as the categorical structure and represented by the paired expressions of eq. 3 & 4, or eq. 5 & 6. Then, we have noted that the integral processing relates to eq. 6 and the separable processing, to eq. 5. Assuming this is the case, it is possible to relates the dominant processing to eq. 5. That is, if \( i \) in \( b_i \) of eq. 5 is infinite (\( \infty \)), we get a supremum and in the process of this mental computation the dominant processing is in operation. It may be recalled that operator \( \sum \) stands for infinite operations, just in case \( i = \infty \). In this sense, eq. 5 is close to the intrinsic nature of the dominant processing and its outcome ‘supremum’. On the other hand, the separable processing is bound to be finite. This is because it is more natural to think that some relevant features available in a specific context (of speech) are bound to be finite. This cognitive restriction may be more apparent in the case of the separable processing; the subjects must be capable of paying separate attention to all relevant
features, on his adopting a separable processing. In any case, if there are infinite things available to him, he would not be able to attend separably and would find a separable processing practically impossible. Further, we have recalled the discussion in §9.2 in which some conceptual affinity between the asymmetric separable and dominant processings is noted. The present discussion supports this impression. Thus, we may argue that eq. 5 accommodates the two processings 'dominant & separable'. The difference between them is that the separable processing refers to finite cases, while the dominant processing refers to infinite cases. Since eq. 6 accounts for the integral processing and eq. 5, both dominant and separable processings, the paired equations of 5 & 6 accommodates the three processings and represents the categorical structure itself. This point seems to explain the tenet of Hirsh & Smale’s or Kobayashi’s proofs mentioned above.

We have noted that because of the MINISSA program, it was not possible to raise the value of r more than 12.0. However, it appears to be possible to get some idea of what it would be, if we could raise the value more than 12.0. The method is based on the following statement made by Arnold. Arnold states that "there exists a one-to-one correspondence between two dimensional solutions in city block and maximum component spaces" and that the two kinds of distance have the following relationships.

\[(9) \quad d'_{ij} = \cos (45^\circ) \quad d_{ij}\]
Arnold, (1971:357)

(where \(d_{ij}\) represents distances in a 2-dimensional city-block space and \(d'_{ij}\) represents distance in a 2-dimensional maximum component space); see quotation below in point (3).

This equality between the city-block and maximum component distances in a two-dimensional space may be regarded as another evidence of equivalence. This time, we examine the logical notion of equivalence in terms of derived distances rather than stress values. We take KOU2 and
Using eq. 9, we calculate predictable maximum component distances, assuming the city-block distances as given. One can compare the observed maximum component distances with the predicted distances.

(.....see Tables 1 & 2 in 6.2.4, Appendices)

Tables in 6.2.4, Appendices compares the observed distances with the predicted distances. When the city-block distances are 0.0000, the predictable maximum distances are not calculated, because such calculations are meaningless. Tables indicates that as r increases, the observed distances approach the predicted distances accordingly. (Although any prediction is not made in the other data, the similar tendency may be easily understood.) This suggests that if we should achieve a sufficiently high 'maximum' for r, the observed distances would have matched with the predicted distances. Assuming that this is the case, we get the equality between the city-block distances and the maximum distances. Since we know already that the city-block distances are all equal to the Euclidean distances in all 10 sets of data. Then, the three kinds of distances will be all equal, if a computer programme manages to accept the 'ideal' supremum metric. The statement is merely hypothetical, but it is worth noting that Arnold's equation helps us to recognize the equivalence between the city-block and maximum distances.

In the above, we have reexamined the characterization of categorical processing (§9.2) in view of the data obtained. This examination suggests that categorical processing relates to integral, separable and dominant processings simultaneously. In this sense, categoriccal processing appears to require the three processings to operate in parallel. The data obtained suggests that the integral and separable processings indeed operated in parallel, since the stress values were consistently identical in all sets of data, bearing clear proof of the categorical processing. But the dominant assumption did not yield
the identical stress with the exception of YAK1, although the categorical processing predict that stress should be identical. Then, this mismatch of stress becomes more enigmatic. This led us to examine the dominant assumption more closely. For this purpose, we used Arnold's equation. The result showed that a sufficiently high r would yield the identical distances to those of integral and separable assumptions, allowing us to conclude tentatively that the subjects are most likely to have adopted the categorical processing in making their similarity judgements.

The performance vs competence distinction into MDS results:
The following brief discussion is motivated by the need to understand more fully why the assumption of dominant processing yielded the slightly superior stress in the 12 sets of data (see Table 5). This superiority is certainly negligible, because the stress values are only better in the third decimal places or even much lower values. Yet this negligible superiority is consistent in that 12 out of the 20 sets of data (age x verb x metric) show this superiority (see Tables 5 & 8). It may not be appropriate to ignore this consistency completely. Our discussion also relates to the above finding that the subjects appear to have adopted the categorical processing in their judgemental process of similarity. By connecting the distinction between performance and competence with the present issue, we will show that the major finding of this section is not contradictory to the slight superiority of the dominant assumption.

At this point, it is useful to recall Diagram 1 in this chapter. The diagram emphasizes the intervention of processing which is governed by such performance factors as the communication pressure of speedy processing. While the prototypical situation relating to competence by definition elicits what is accessible in one's memory and renders these accessible entities available (probably) in one's immediate memory. But, under usual circumstances one's entire knowledge of language, even of one lexis, may not be made available in full. (It may be recalled
that the present distinction between competence and performance is simple (see chap. 3 and this chapter: i.e., all vs some). According to Diagram 1, the MDS representation is the product of a processing which the subject adopts at the time of making judgements. As we have seen in § 9.1, it has been experimentally proved that a specific choice of processing is governed by performance factors such as the nature of an experimental task, the structure of stimuli and the level of cognitive maturation. Assuming that a special choice is made by the subject on the basis of such performance factors, we may suppose that he chooses one type of processing rather than the others, on the ground that it is most relevant for the purpose of meeting his communicative needs. That is, there must be one predominantly operative processing at any given level of performance. This view may account for the superiority of stress values of the dominant processing over the integral and separable processings. Of the 20 sets of data (verb type × age × dimensionality), 12 sets of data show the superior stress in the dominant processing.

KOC3  KOK1  KOK2  KOU1  KOU2  YAK1  YAK2  YAU2
2/3    3     2/3    2     2/3    3     2/3    2

Table 8
(Table 8 is estimated on the assumption that the subject chooses one type of processing rather than two or more.)

The above argument suggests that the superior stress of the dominant assumption is a reflection of performance factors. However, it is in theory possible to assume that the subject can perform several types of processing in parallel. As we have seen above, in the case of the categorical processing it is highly likely that the subject manipulates several processings simultaneously. This simultaneous multiple processing agrees with the present characterization of prototypical situation and hence the present definition of competence in word-meaning. In this sense, the prototypical situation may represent the cases in which several relevant types of processing are
triggered in parallel.

Thus, we may say that due to some performance factors, one kind of processing may be singled out by the subject; and MDS may highlight his choice. On the other hand, competence involves several processings at the same time. As the NL structure revealed in performance is a part of competence, so is a specific processing used by the subject for the purpose of achieving his need. A specific form of processing at the level of performance is one of several processings available in competence. Thus, that one type of processing is predominant is not a problem here. In this sense, the slight superior stress yielded by the dominant assumption is not a problem, once we view it as reflecting some performance factors. Thus, the present tentative conclusion that the present data essentially conforms to categorical processing appears to be valid.

The above tentative conclusion may have the following implication. As we have seen in §9.2, Shepard considers the categorical structure as unsuitable for the dimensional scaling of MDS. The present result also suggests that the MDS representation is not effective in revealing the categorical structure, because we need a representation which combines information from at least three metrics. We must remember that, as Shepard (1974) says, the Euclidean metric is robust; in this sense, the Euclidean metric may be sufficient in producing a fair approximation which takes the place of the other metrics -- this view is supported by the empirical investigation of the present study, since the other metrics yielded 'equivalent' stress and highly similar configurations. But, when the accuracy of the derived distances is at issue, as in Chap. 4 and 10, the Euclidean metric cannot take the place of the other metrics.

The above view that MDS may not be effective in rendering the categorical structure may appear to be contradictory to another observation made of Table 8, since Table 8 states that MDS is effective in representing performance aspects of data. The argument is consistent in that all 20 sets
of data conform to the categorical structure/processing which Shepard regards as unsuitable for the dimensional scaling of MDS. MDS is not effective in this sense. But we must bear in mind that the MDS method manages to reveal that the present data ought to be represented non-dimensionally and to yield the equivalent stress values from the three different metrics. Because this sort of technical capacity is implemented in MDS, we would argue that it is significant for the supremum (in practice quasi-maximum) to have yielded consistently slightly better fit to the data, although the differences with the other two metrics are very small. And this superior stress is, as we have seen in §9.1 and above, the result of performance factors including a specific choice of processing the subject makes. For the same reason, the MDS configuration is useful in revealing the state of NL which is affected by performance factors. But, MDS is unlikely to reveal competence as it is. If competence is categorical, it will be distinct from the MDS representation. For this reason, when the derived distances need to be accurate, i.e., if the investigator is interested in basing his experimental predictions upon these derived distances, the MDS results do not appear to be useful.

The above discussion suggests the following. (1) It is most probable that the subjects have adopted a categorical processing and that the NL interlexical structure is best treated as categorical. For this reason the MDS distances may not be used effectively for our experimental predictions in chap. 10. (2) On the other hand, the MDS result is effective in revealing which is the most predominant form of processing. (3) For a proper representation of the categorical structure, it appears to require a procedure which allows for some parallel operation of the various metrics to accommodate eq. 3 & 4 or eq. 5 & 6. Since this is not made possible in MDS, it is thus inevitable that the MDS analysis is restricted to the performance side. This MDS's emphasis on the performance side appears to be reflected in the superior stress yielded by the
dominant processing in the data.

§ 9.3.2 (3) specifying a form of processing based on unit circles

We have seen above that the MDS results may relate more intimately to performance than to competence. In this connection, it is interesting to see that YAC3, YAK1 and KOU2 turned out to be highly degenerated structures which resemble the unit circles of the supremum (maximum) metric and of the city-block metric respectively. According to Figures in 6.1, Appendices, the configurations of YAC3, YAK1 and YAU2 in the 2-dimensional spaces look 'like' (oblong) squares respectively (see below). The configuration of KOU2 is diamond-shaped. The stress values are excellent in these cases. This suggests that the apparent "degenerated" configurations are the outcome of the 'normal end of runs.' In this sense, we may regard the configurations as reflecting faithfully some performance factors under which the subjects judged the extent of similarity among all pairs. If so, the respective configurations indicate clearly that the subjects belonging to groups YAC3, YAK1 and YAU2 performed a dominant processing, and those in KOU2, an (asymmetric) separable processing.

(.....see 6.1, Appendices)

In YAC3 and YAK1, the configurations in the 2-dimensional city-block (and Euclidean) spaces represents oblong squares, 4 corners of the square consisting of stimulus points 8, 5, 7 & 10 in YAC3, and 6, 9, 8 & 11 in YAK1. Both Arnold(1971) and Shepard(1974) write that the unit circle in the 2-dimensional maximum component space is a square (see §9.2). The shape in YAC3 and YAK1 is not a square in the city-block and Euclidean spaces. But, Arnold indicates elsewhere that the unit circle of the 'dominance' metric becomes parallelograms when it is deliberately plotted in the Euclidean space (1971:361).
The configurations in YAC3 and YAK1 conform with parallelograms in the Euclidean and city-block spaces. This suggests that the subjects belonging to YAC3 and YAK1 may have adopted a dominant processing. But the respective plots in the maximum component space (r=12.0) shows that the above 4 stimulus points are converged into two; in YAC3, point 5 is overlayed with point 8 and 7, with 10; in YAK1, 6 is overlayed with 9 and 8, with 11. These configurations is not the shape of rectangle, consisting of two points. This may imply that the configurations in YAC3 and YAK1 does not stand for the unit circle of the dominance metrric. However, the printing technique ( either of line printers or of the MINISSA program) may be questioned for the following reason.

<table>
<thead>
<tr>
<th>R</th>
<th>Pairs</th>
<th>YAC3</th>
<th>YAK1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>8/5</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2.0</td>
<td>10/7</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>12.0</td>
<td>6/9</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>8/11</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 9 interpoint distances

We will take YAC3 first. when r = 1.0 or 2.0, points 5, 8, 7 & 10 are plotted in the 2-dimensional space, forming the four corners of the rectangle. But, as Table 9 shows, the distances between 8 and 5, and between 10 and 7 are 0.0000. On the other hand, in the maximum component space, the distances between these pairs are greater than those in the city-block or Euclidean spaces. In spite of this, in the maximum component space 5 is overlayed with 8 and 7, with 10. This is unsatisfactory. The method of printing does not appear to be sensitive to the minor differences. While in the city-block or Euclidean spaces the interpoint distances of 0.0000 are printed apart, then, the configuration in the maximum component space ought to have been represented as a rectangle at least.

In the case of YAK1, the interpoint distances between two pairs (6 & 9, and 8 & 11) are equally 0.0000 in the three
kinds of metric spaces. However, points 6, 9, 8 & 11 are plotted distinctly, consisting the four corners of the rectangle in the city-block and Euclidean spaces, but this is not the case in the maximum component space. This also illustrates a minor defect in the MDS printing technique.

We have indicated three points above. (a) Although YAC3 and YAK1 do not conform to the shape of the unit circle of the dominance metric (i.e., square), the respective configurations are rectangles which correspond to the unit circle of the dominance metric plotted in the Euclidean space. (b) There is a minor defect in the printing technique. In addition to this, we recall that the current MINISSA does not accept \( r \) if it exceeds 12.0. It is possible to think that if we have achieved some ideal \( r \) to satisfy some requirements to call \( r \) 'supremum', we might have obtained the configuration of square which predicts the dominant processing legitimately. For these reasons, we may conclude tentatively that YAC3 and YAK1 are likely to have followed the dominant processing.

The configuration in KOU2 is interesting. It is distinctly diamond-shaped, suggesting that the subjects belonging to this group have adopted the (asymmetric) separable processing. We cannot determine whether it is an asymmetric separable or separable processing. This topic is beyond the scope of the present research.

YAC3, YAK1 and KOU2 (including YAU2 to some extent) are also interesting in that the following observation made by Arnold (1971) is born out fairly accurately.

Equation 5 represents the distance between two concepts, the \( i^{th} \) and \( j^{th} \), in two-dimensional city block space,

\[
d_{ij} = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}|,
\]

where, as in Equation 4, the \( x \)s are coordinates for the concepts on the reference axes. Consider next what happens when the configuration is rotated 45° with respect to its reference axes and the distance,
d' is again calculated, this time using the maximum component distance formula,

\[ d'_{ij} = \max \left( \cos(45°) |(x_{i1} - x_{j1}) + (x_{i2} - x_{j2})|, \cos(45°) |(x_{i2} - x_{j2}) - (x_{i1} - x_{j1})| \right) \]  

Now it is clear that if the pair of binominals that appear in both of the terms between the brackets in Equation 6 have the same sign, the term to the left of the comma will be the maximum. If the signs are different, the term to the right will be the maximum. In either case we may write

\[ d'_{ij} = \cos(45°)d_{ij} \]  

In short, there exists a one-to-one correspondence between two-dimensional solutions in city block and maximum component spaces. Arnold, (1971:357)

Arnold indicates two things. (a) There is an isomorphic correspondence between the city-block and maximum component distances. (b) This isomorphic correspondence is apparent when "the configuration is rotated 45° with respect to its reference axes." These two points appear to be born out by the present data. The first point has been examined in (2) of this section, where for the purpose of illustration YAC3 and KOU2 are taken and ascertained that Arnold's eq. 5 appears to account for the present data. As for Arnold's second point (b), we can compare the configuration in the city-block space with that in the maximum component space (see 6.1, Appendices). As the following schematic diagrams depict, the configurations in the city-block spaces are rotated roughly 45° in the maximum component spaces; the reference axes are indicated by the dotted lines below:
Since Arnold's computational prediction is born out fairly well, we may regard the configurations of YAC3, YAK1, YAU2 and KOU2 as the outcome of the normal runs of MINISSA rather than as highly "degenerated" results. This also supports the above tentative conclusion that the subjects in YAC3, YAK1 and YAU2 have adopted a dominant processing in their judgemental process of similarity, and KOU2, an (asymmetric) separable processing. As we have argued in (2), these findings relate more closely to the performance aspect of processing than to the competence aspects of judgemental process. Since we have specified (as a reflection of performance) the respective forms of processing in YAC3, YAK1, YAU2 and KOU2, Table 8 above needs to be modified accordingly. Table 10 includes these modifications.

<table>
<thead>
<tr>
<th>KOC3</th>
<th>KOK1</th>
<th>KOK2</th>
<th>KOU1</th>
<th>YAC3</th>
<th>YAK1</th>
<th>YAK2</th>
<th>YAU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/3</td>
<td>3</td>
<td>2/3</td>
<td>2</td>
<td>2/3</td>
<td>3</td>
<td>2/3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 10

§ 9.3.2 (4) developmental trends seen in processings

Developmental pattern is discussed here. First, the development is studied in terms of the ranked SVs which were calculated by the Thurstone-Torgerson method and used as input to MINISSA. Second, the developmental pattern of processings is viewed in relation to the MDS configurations.

We used as input to MINISSA the ranked SVs. These similarity measures were based upon the SIM (see F-matrix, P-matrix, X-matrix, SVs in 7.1.1, 7.1.2, 7.1.3 & 7.1.4, Appendices and the 3 kinds of matrix are explained in chap. 7). Since the derived SVs were ranked in order to be used as input, Spearman's rank-order correlation coefficients were calculated in a pairwise manner to see whether the ranked SVs show any significant relationship among the 10 sets of data (age x verb). The results are presented in 6.2.7, Appendices.
Apart from YAU2, the ranked SVs among the other groups correlates with each other significantly. It is surprising to remember that the MDS configurations themselves differ markedly, although there are significant correlations among the input data obtained from the four groups of subjects.

The 10 sets of SVs were studied in terms of regression lines and scattergrams, in order to see how well SVs in each group can predict the SVs which were obtained by the one-year older group. It may be recalled that the subjects were grouped according to their school year and the difference in age was one year. (Unfortunately, there were no subjects who were in the 3rd year of high school, since they were unable to take part in any of the present experiments, due to the preparation for their university entrance examinations.) Tables and Figures in 6.2.8, Appendices list the results of regression lines and scattergrams. The analysis of the ranked SVs in terms of regression lines and scattergrams is useful for the present purpose. For instance, the regression of KOK1 on KOC3 represents the prediction of KOK1 (the group one year older than KOC3) on the basis of the performance made by KOC3. The scattergram can show the amount of prediction made by KOC3 by the ratio of dispersion. The scattergrams in Appendices do not specify the amount of dispersion numerically, because SPSS was used here. In this sense, we can only estimate the amount of dispersion in an impressionistic fashion, by looking at the scattergrams. But, both the regression analysis and the scattergrams were made pairwise. This enables us to study the data as the groups develop, comparing at one-year, two-year, three-year, four-year and five-year intervals. The pairwise scattergrams show that the greater interval there is, the greater amount of dispersion there is. This suggests that the greater the interval is, the lesser the prediction of regression becomes. Thus, the present regression analyses coupled with the scattergrams indicate the steady growth of SVs among the five groups of subjects participating.
It was suggested above that in spite of the diverse visual differences in the MDS interlexical structure representations, there was some orderly progression of SVs which were used as input to MINISSA. This orderly development may relate to the progression of NL interlexical structures from differentiation to unification in kowasu and the reversed pattern in yaburu. This is discussed below.

The following schematic representations of the MDS results suggest, the development of the NL interlexical structures indicates the pattern from differentiation to unification in kowasu and the pattern from unification via differentiation to unification in yaburu.

The pattern of development in kowasu is reversed in yaburu. The progression from differentiation to unification in kowasu and the reversed pattern in yaburu is also depicted by Shepard's diagrams in 6.1, Appendices. Shepard's diagrams of YAC3, YAK1 and KOU2 represent "step functions" which are reflected as highly converged results in the configurations (see the 2-dimensional spaces). These data appear to reveal a high degree of unification. On the other hands, Shepard's diagrams of the other sets of data show a greater amount of dispersion, reflecting the degree of differentiation.
The term unification is applied to YAC3, YAK1 and KOU2, since several stimulus points are unified into one (see 6.1, Appendices). Whereas in the other 7 sets of data, the respective stimulus points are well differentiated in the MDS configurations. These terms are chosen, because they are descriptive of what an outside observer would have used to express his impressions of such MDS configurations in 2-dimensional spaces. Let us assume for a moment that the MDS results faithfully represent the NL interlexical structure which a language-learner possesses. From the learner's point of view, unification may mean something different from the observer's usual notion. Unification may imply that a given semantic feature which is represented as a dimension in the MDS results stands for a highly general concept in his mind. Or alternatively, it may mean the presence or absence of a given semantic feature. The first interpretation is possible, since unless a feature is general, several words cannot be represented as converging into one point, and the feature must be applicable to such words. The second interpretation is also possible, since the presence of a given feature may be represented as one discrete point, or else it may be regarded as a non-member of this discrete point. There is one point in common with these interpretations. That is, the learner knows that the same feature exists in all words which are represented as one point in his internal NL structure. Therefore, there is no need for them to be differentiated in his semantic map. For this reason, unification appears to presuppose that a given semantic feature is general. We can say the opposite of general in the case of differentiation. The words which are shifting towards differentiation contain specialized semantic features. In this sense, the processes of unification or differentiation indicate a qualitative change in semantic features from general to special. This internal change is examined in chap. 10 in relation to the notion of "representational reorganization" which is revealed as the "U-shaped behavioural growth curves": see Strauss.
We have seen above that YAU2 did not correlate significantly with the other 4 sets of data in terms of the ranked SVs which were used as input to MINISSA. YAU2 was the only group which showed very poor rank-order correlation coefficients with the rest of data involving yaburu. One can account for this peculiarity in connection with the unification and differentiation processes. (A) Diagram 3 indicates that the order of occurrence of these processes is reversed in yaburu. Yaburu progresses from unification to differentiation returning to unification in YAU2, while kowasu develops from differentiation to unification. (B) In addition to this reverse order in yaburu, we notice that the extent of differentiation is similar to that of KOC3, the youngest group of all dealing with kowasu. (C) Furthermore, yaburu appears to be less sophisticated or solid in terms of the internal structure and the form of processing the subjects appear to have adopted. YAU2 was the oldest group of all involving yaburu, but the degree of differentiation is similar to the youngest group in kowasu. YAC3 and YAK1 appear to indicate a dominant processing, but very poorly (see (2) above). On the other hand, in the case of kowasu, we have seen that the oldest group appears to have processed a set of items separably, adopting either an asymmetric separable or a separable processing. This is revealed distinctly as the corresponding unit circle of city-block metric. Since in kowasu the ranked SVs in the respective groups indicate a steady growth and these SVs were used as input to MINISSA, it is therefore reasonable to regard the development of processings as steady. Besides, the examination of stress values above did not indicate any radical change among the five groups of subjects. These observations suggest that the five groups of subjects used the same type of processing as KOU2, in coping with the performance aspects of the judgemental process. There is a strong sense of solid steadiness in kowasu, which is lacking in yaburu.

What has emerged in the three observations (A), (B) & (C)
is a possibility that YAU2 marks the turning point moving towards the progression pattern of kowasu. This may account for the reason why the only YAU2 showed very poor correlations with the other 4 groups involving yaburu.

§ 9.3.3 conclusion

The following findings and topics of discussion were covered in this chapter.

1. MDS can be used to specify what form of processings is adopted by the subjects at the time of making similarity judgements. The previous views concerning the correspondence between metrics and processings are summarized in § 9.2 (see Table 2): integral processing (Euclid), separable processing (city-block), asymmetric separable processing (city-block) and dominant processing (supremum). The method of specification involves the adjustment of parameter r (Minkowski's general metric r), the examination of stress (index representing the amount of discrepancy between input and output) and unit circles (isosimilarity contour). MDS computes the latter two.

2. We have tentatively characterized categorical processing based upon Shepard (1974). The present characterization accords with the present definition of word meaning (§ 9.2 & § 9.3.2 (2)).

3. There has not been suggested any underlying metrics which predicts categorical processing. The two proofs mentioned in § 9.2 indicate that the categorical structure involves the logical notion of "equivalence": that is, whatever norms (metrics) we supply, as a parameter of r, the results of computation would be equivalent. This suggested that the various values of r would yield equivalent distances and therefore the same stress. This version of "equivalence" was regarded as the indication by which we can identify categorical processing.
4. The above logical notion of "equivalence" was then translated into the two kinds of practical arithmetics (a) & (b). These two were not satisfactory in view of the built-in mechanisms of MDS, e.g., iterations and rotations (see above).

5. The observations made in points 3 & 4 led us to examine the nature of categorical structure which is defined by the paired equations of 5 & 6 or of 3 & 4. Garner assumes that the structure of stimulus (or a set of stimuli) corresponds to the form of processing in theory. If this is the case, the categorical structure needs to be processed categorically. The examination of the paired equations of 5 & 6 suggested that the paired equations should accommodate the four processings simultaneously. This implies that categorical processing involves the parallel operation of the four processings. The parallel operation may mean one is part of the others simultaneously, which implies in turn one is equivalent of any one of them. The above definition of equivalence is thus justified, and, accordingly the operational definition of equivalence (i.e., the equivalent distances and stress (see point 3)) gains credence, but it may only be one apparent and obvious indication of equivalence among all other possibilities.

6. The operational definition of equivalence which predicts categorical processing was largely born out by the data. Both integral processing ($r=2.0$) and separable or asymmetric separable processing ($r=1.0$) yielded identical stress and distances among all subjects involving two verbs (kowasu & yaburu). The only exception was the dominant assumption. The assumption of dominant processing yielded slightly superior stress (in the 4th decimal places) to the other three assumptions. This amount of superiority may be negligible. In addition, MINISSA (the computer program used) does not accept the value of $r$, if it exceeds 12.0. This value is not a maximum or supremum ($r=\infty$) which identifies this dominant processing. Because of this practical unfeasability, it appeared that we did fail to
demonstrate the equivalent distances and stress. But it appeared that if we increased the value of $r$ sufficiently, we could have obtained equivalence. For this reason, we calculated some predictable distances on the basis of Arnold's equation. The result showed that as $r$ increased, the predicted distances approached the Euclid and city-block distances which were identical to each other in the MDS results. For these reasons, the dominant processing could have yielded the distances (and therefore stress) identical to the distances under the other processing assumptions.

7. Thus, the ten sets of data (age $\times$ verb) appeared to support the categorical assumption. That is, all subjects appeared to have intuited the NL interlexical structure as categorical, processing a set of NL items categorically. As Shepard (1974) makes clear, the dimensional scaling of MDS is not appropriate for any categorical structure. This suggests that the MDS representations of NL structure may not be relevant for the present purpose of explaining the translatability judgement on the basis of the MDS representation of NL structure. However, as the free recall experiment (chap. 6) indicated, the clusters in the MDS configurations are psychologically meaningful. In this sense, we will make use of the MDS configurations as our guiding maps --- that is, we do not use the MDS distances to examine the translatability judgement data in chap. 10. The experimental results in chap. 7 showed that the Thurstone-Torgerson method was promising as the measurement of similarity. For this reason, the SVs on the basis of SIM will be used in chap. 10.

8. The discussion in § 9.1 & § 9.3.2 (3) suggests that the MDS results might be more closely related to performance. It was argued that at the level of performance the subject would rely on one type of processing rather than the others, largely due to such performance factors as the nature of the structure of stimuli or specific situational demands (e.g., speedy processing). This
selection on the part of the subject would be reflected on the MDS results. This appeared to account for the slightly superior stress under the dominant assumption. The reliance upon this dominant processing was most apparent in YAC3 and YAK1 whose configurations resembled the unit circle of the dominance metrics (see §9.3.2 (3)). As a reflection of performance, the MDS results indicated that yaburu tended to elicit the dominant processing more than kowasu did. On the whole, yaburu appears to elicit less advanced structural understanding on the part of the subjects. It is noteworthy that KOU2 indicated the (asymmetric) separable processing because of the distinct diamond shape corresponding to the separable processing. The MDS configurations of kowasu suggested on the whole there was the sense of homogeneity in the structural understanding exhibited by the MDS configurations. In this sense, the five groups of subjects might have adopted equally the same (asymmetric) separable processing, coping with the task of making similarity judgements.

9. This pilot study has a few developmental implications. (a) When we looked at the MDS configurations developmentally, kowasu suggested the shift from differentiation to unification; on the other hand, this tendency was reversed in yaburu, i.e., there was a movement from unification via differentiation to unification (see §9.3.2 (4)). (b) The amount of differentiation in YAU2 (the oldest group involving yaburu) was similar to that in KOC3 (the youngest group involving kowasu) (see Diagram 3). These points (a) & (b) seem to suggest that the lexical understanding depends largely on the nature of verb and a form of processing adopted. In this sense, the age of the subjects may not be much to do with the depth of structural understanding, when it reflects some communicative pressures imposed by the experimental situation.
Chapter 10: translatability experiments - developmental trends

§ 10.0 introduction

The foregoing chapters (5 -9) dealt with the specific issues which had been brought up by my follow-up study of Kellerman (1977 & 1978). This chapter is intended to provide a more focused analysis than the follow-up study in chap. 4, by making use of some of the relevant findings which were presented in the previous chapters.

The experiment among the youngest group (coded as YAC3 and KOC3) was administered on the same day when the subjects learned the verb 'break' as a part of their regular lesson. In terms of the 'quality' of 'experimental' subjects, they were to all intents and purposes ideal subjects. They were most likely of all the subjects participating to be free of extraneous factors such as long-term confusion of TL and NL concepts. For this reason, KOC3 and YAC3 are discussed separately from the older groups in §10.1. The older subjects are discussed in § 10.2 (kowasu) and in §10.3 (yaburu). The data gathered from these older subjects are analysed in terms of developmental trends as well as the effects of processing upon the judgemental data. The specific hypotheses in each section are presented in correspondence with the section organization.

§ 10.1 experiment 1

This section is divided into four parts. §10.1.1 deals with experimental design. §10.1.2 explains the method of administration. §10.1.3 gives hypotheses. §10.1.4 discusses the results obtained.

§ 10.1.1 experimental design

This section deals with KOC3 and YAC3. The method of administration follows Pilot study 1 (chap.4), except for one instruction and the number of trials. The subjects
in KOC3 and YAC3 were told that one of the items on the list was translatable into English, using the English verb 'break'. In the case of \textit{kowasu}, the translatable item was item 1 (mokei: scale model). Since this is one of the concrete items (i.e., the object nouns are concrete nouns), this provides us with 'concrete' condition. In \textit{yaburu}, the translatable item was item 2 (houritsu: law). Since this is one of the abstract items, it provides us with 'abstract' condition. The purpose of having these two experimental conditions is (a) to examine the dominance effect discussed in chap. 6. This dominance effect relates to Rosch's basic level, since this basic level tends to be concrete (see chap. 3 and appendix 2). (b) As pilot study 1 indicated, \textit{yaburu} showed the process of concretization which was peculiar to this verb but the concretization was not observed in \textit{kowasu}. This difference was explained by the syllabus design. In the standard textbooks which the subjects studied, the association between \textit{kowasu} and 'break' was introduced with concrete nouns (see appendix 4). They learned the literal use first. It was likely that the TL sign 'break' was associated with the concrete NL meaning. Whereas in the case of \textit{yaburu}, before they learned the literal use of 'tear', they learned the metaphorical use of 'break' (i.e., breaking one's promise). The metaphorical use of 'break', as in the example, translates into the Japanese word \textit{yaburu} which stands for the English 'tear'. This appeared to establish the confusion of the NL concept (\textit{yaburu}) with the TL sign 'break'. This may be regarded as an 'over-generalization error'. But the process in which the error was produced appeared to be that the abstract NL concept was gradually consolidated as a concrete mental entity in relation to the TL sign. In any case, the difference in the syllabus design initiated the peculiar confusion of TL and NL concepts in \textit{yaburu} which was called concretization in chap. 4. The present experiment simulates this effect of syllabus design.

In Pilot study 1, the number of trials was three. This number was chosen because of the three factors. (a) The
probability of a 'sure' item to be included as the member of consistent category should correspond to (b) the probability of excluding random answers, which enables us to remove 'consistent' answers by chance, so that (c) the interlanguage competence was distinguished from the variable IL performance (see §4.3.3). This consistency criterion (realized as the three independent trials) was effective for our purpose, particularly because the subjects who took part in the translatability experiment were 'fossilized' IL speakers who had studied English in Japan and acquired English in Edinburgh for a long period. Because the extent of fossilization was expected to be sufficient, the consistency criterion was effective among these subjects.

On the other hand, the subjects in KOC3 and YAC3 have learned English for only two years. This period of exposure to TL may suggest an intermediate level by European standards, but as the examination of school textbook examples suggest (appendix 4 & chap.4), a two-year study period in a Japanese state school would mean they were practically total beginners, especially at the time when this experiment was run. More important, they had just learned 'break' on the day of the present experiment. With respect to this vocabulary item, they were absolute beginners. This suggests that a greater amount of variability was likely than was the case with the subjects in Pilot study 1. In this experiment, we make use of SIM. As Atkinson & Juola's experiment makes clear (chap.7), it is important for us to obtain some asymptotic (flat) performance from the subjects. Otherwise, SIM may not be as effective as it should be. For this reason, the number of trials must be increased. But as I was only allowed to spend one school session (45 minutes), it was decided that the number of trials would be five.

§ 10.1.2 administration

One hundred 3rd year students at Shonan Middle School took part in the experiment. Half of the subjects were present-
ed with a set of 11 kowasu sentences and the other half, with a set of 10 yaburu sentences (see chap. 5 & 6). The former group was coded as KOC3 and the latter, YAC3. They learned that one of the sentences could be said with the English verb 'break'. In KOC3, they learned the 'break' sentence with the concrete object noun mokei (scale model). YAC3 learned the 'break' sentence with the abstract noun houritsu (law). Then, both groups were asked to guess whether the remainder of the NL list could be said with the same English verb.

The subjects were allowed to spend 30 seconds per sentence. The same procedure was repeated five times with distractor tasks intervening. The distractor task consisted of mental arithmetic and category naming. In the former task, the subjects were asked to add up whole numbers which were orally presented for five minutes. In the latter task, the subjects were instructed to write down words that could be included in the categories of western dishes, fish, oriental dishes, and animals with four legs (five minutes). Excluding the time it took to distribute pamphlets, about 45 minutes were used to complete the task.

§ 10.1.3. hypotheses

There are four research interests in experiment 1.

(1) This experiment is designed to test the dominance effect (see §10.1.1). KOC3 learned the association of the English sign 'break' with the NL concrete concept. This provides the concrete condition. YAC3 learned the association with the NL abstract concept, representing the abstract condition. The dominance effect predicts that KOC3 will learn the association better than YAC3.

Rosch's basic level predicts the same result as the dominance effect does. As we have seen in chap.3, the basic level relates to learning. Words at the basic level are
easier to learn than those at the higher or lower level. According to Rosch (see chap.3), a basic level is defined as the most inclusive member of a category which has the maximum within-category similarity. This means in the present context that a word at the basic level has most of its semantic features in common with the other words belonging to the same category. Kowasu and yaburu belongs to the verbs of 'destruction'. Table 1 (in appendix 3) lists this category of 'destructive' verbs in Japanese. The listing is not complete, but it is sufficient for our purpose. On the basis of Table 1, it is reasonable to say that kowasu has the greatest number of features part of which are shared by other verbs. In this sense, kowasu is at a basic level. For this reason, we predict that the association of the NL concept (kowasu) with the TL 'break' will be learned better than the association of the yaburu concept with the TL 'break'.

(2) Based upon the experimental results in chap.9, we make use of SIM rather than MDS for the measurement of similarity. SIM entails that the judgemental process involves two-item comparison. But the experiments in chap.6 showed that this paired comparison appeared to play some major roles in the process of translatability judgement. In this sense, we have justification in using SIM. Among other things, the major merit of SIM was discussed in §7.4. That is, SIM can determine the judgemental limit (see chap.3 & 7). We have demonstrated experimentally that the judgemental limit can operate inductively and deductively (chap.7, particularly §7.4). A learner induces for the first time the range of application of a TL sign with the corresponding NL concept in mind. This initial experience establishes the inductive limit on the part of the learner. (In trial 1 in the present experiment the subjects would have arrived at this judgemental limit.) The same learner may deduce the translatability of a novel NL concept, using the judgemental limit which he has acquired previously. This illustrates the function of judgemental limit, when it is used deductively. In this experiment we will first
of all examine the plausibility of judgemental limit.

(3) Chap.3 discusses chained reasoning among L1 children (see §3.4). Since expriment 1 involves five trials, chained reasoning is likely to occur in these successive trials. Once we estimate the respective inductive limits of kowasu and yaburu on the basis of the data obtained, we can find out whether or not the projective limit can identify chained reasoning.

(4) The last research interest is to examine the effect of data processing upon translatability judgement. In chap.9, we have seen that YAC3 appears to have used domi-processing at the level of performance. On the other hand, KOC3 appears to have used (asymmetric) separable processing. In §9.3.2(2) we argued that the 'ideal' cases of dominant and separable processings would require both eq. 5 and 6. But, here, we assume that the dominant processing (of the kind that YAC3 demonstrated) may be more closely related to eq. 5. On the other hand, we assume that KOC3 relates to both eq. 5 and 6. These assumptions are reasonable in that the dominant processing exhibited by YAC3 resulted in a rather poorly structured configuration, while the separable processing (exhibited by KOC3) produced a properly structured MDS representation (see 6.1 Appendices).

Arable and Boorman (1973) and Shepard (1974) discusses the MDS method in terms of lattices (see §9.2). As we have seen in chap.9, eq. 5 and 6 are properly regarded as belonging to lattices or Boolean algebras. The following basic knowledge of lattices follows Matsumoto (1980: chap.1). There are two kinds of lattices: "modular" lattices and "distributive" lattices. The inferior type of dominant processing relates to the former modular lattices. Modular lattices are inferior to distributive lattices, since they do not satisfy 'distributive law'. Distributive lattices exhibit more 'well-defined'
structures. The simple contrast between modular and
distributive lattices at least case our intuition about
the MDS configurations yielded by YAC3 and KOC3
respectively.

We have assumed above that YAC3 relates to modular lattices
and that KOC3 relates to distributive lattices. We regard
these assumptions as reflecting the different processes
adopted by YAC3 and KOC3 respectively. We can make one
interesting prediction concerning the 'sure' item. Matsu-
moto's theorem (1.3.5) and theorem (1.3.6) prove that there
are more than one 'relative complement' in modular lattices
and that there is only one relative complement in distribu-
tive lattices (Matumoto's, (1980: 13-17)). This notion
of relative complement is useful in the judgemental proc-
ess of translatability. The subject compares the 'sure'
item with the other items on the list, questioning how
similar each item is to the 'sure' item. The subject looks
at the other items in contrast with the 'sure' item. This
implies that the other items are in the complementary domain
of the 'sure' item. In this sense, all the other items
are possible candidates for a 'complement' relative to
the 'sure' item. Accordingly we will have more than one
complement in YAC3, but there will be only one complement
in KOC3. We will look at how this relative complement
works in later trials.

§ 10.1.4. results and discussion

(1)
The dominance effect (the superior effect of the concrete
condition to the abstract condition) is discussed here.
While all subjects remembered in the later trials what
the initial association was for kowasu, some subjects in
YAC3 appeared to have lost their memory about what they
learned in trial 1. Table 1 indicates this difference.
Table 1

In §10.1.3, we regarded the initial stimulus in kowasu as at a basic level, while in yaburu the initial stimulus is not at a basic level. In this sense, the result shown in Table 1 was expected. Table 1 also shows that more subjects chose kyotei than the initial item horitsu. The subjects' memory of the stimulus item appears to be more stable for the concrete condition than for the abstract condition. Kibou was included in the two lists. This item is remotest (in terms of SV) from the 'sure' item for both verbs (see 6.2, Appendices). But kibou yielded a greater amount of variability in yaburu than in kowasu:

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T3 - T5</th>
</tr>
</thead>
<tbody>
<tr>
<td>kowasu/mokei</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>yaburu/horitsu</td>
<td>50</td>
<td>41.0</td>
</tr>
<tr>
<td>yaburu/kyotei</td>
<td>42</td>
<td>44.3</td>
</tr>
</tbody>
</table>

Table 2

Tables 1 & 2 suggest that the concrete condition creates the tendency to cause the greater stability in memory and the lesser variability in responses than the abstract condition does. (In this sense, it is important for us, teachers, to use concrete items at the basic levels in foreign language instruction.)

(2) The method of estimating judgemental limit is the same as in chap.7 (see exp.1, §7.4). 7.1, Appendices, presents SVs and DDs derived from SIM. 7.1 also includes F-matrix,
P-matrix and X-matrix. From the experiment, we obtain the proportion of subjects who regard a given item as translatable. Using this information in combination with SV and DD, SIM computes judgemental limits (beyond which the subjects reject an item as untranslatable). Table 3 presents the result yielded at trial 1 with respect to kowasu.

Kowasu (trial 1)

<table>
<thead>
<tr>
<th>choice</th>
<th>SV</th>
<th>DD</th>
<th>J.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mokei</td>
<td>50</td>
<td>1.7076</td>
<td>0.3641</td>
</tr>
<tr>
<td>taicho</td>
<td>3</td>
<td>1.5800</td>
<td>0.5353</td>
</tr>
<tr>
<td>kikaku</td>
<td>16</td>
<td>1.6427</td>
<td>0.5102</td>
</tr>
<tr>
<td>taisei</td>
<td>11</td>
<td>1.9446</td>
<td>0.3425</td>
</tr>
<tr>
<td>kibun</td>
<td>2</td>
<td>2.0232</td>
<td>0.3901</td>
</tr>
<tr>
<td>kibo</td>
<td>4</td>
<td>1.8322</td>
<td>0.3340</td>
</tr>
<tr>
<td>okimono</td>
<td>46</td>
<td>1.4655</td>
<td>0.4743</td>
</tr>
<tr>
<td>kosho</td>
<td>6</td>
<td>0.8997</td>
<td>0.3451</td>
</tr>
<tr>
<td>kata</td>
<td>26</td>
<td>1.8322</td>
<td>0.3340</td>
</tr>
<tr>
<td>tofu</td>
<td>41</td>
<td>1.8052</td>
<td>0.4157</td>
</tr>
</tbody>
</table>

| 0.8328  | 1.1810 | 1.4716 | 1.7648 | 2.1308 |

J.L. : Judgemental Limit

Table 3

The set of judgemental limits shown in Table 3 is derived from the subjects' performance in trial 1. So, we call these judgemental limits 'inductive limits' (see §7.4 & §10.1.3). Table 3 indicates that all the judgemental limits are within the range of two category boundaries $T_2$ and $T_3$. This range corresponds to "some relationship" in the 6-point scale we presented to the subjects (see chap.5). It suggests that when the subjects feel the 'sure' item has at least "some relationship" to an item they are judging, they would regard the item as translatable. The judgemental limit is a minimum requirement. If an individual subject's SV is smaller than a given judgemental limit, he will judge the translatability of an item with greater confidence. If his SV exceeds the limit, he will reject the item as untranslatable.

The data obtained from YAC3 were analysed in the same way
as KOC3. Table 4 presents the results of this calculation.

### Yaburu (trial 1)

<table>
<thead>
<tr>
<th>choice</th>
<th>SV</th>
<th>DD</th>
<th>J.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>chinmoku</td>
<td>10</td>
<td>1.4822</td>
<td>0.5669</td>
</tr>
<tr>
<td>horitsu</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yujo</td>
<td>16</td>
<td>1.1662</td>
<td>0.8083</td>
</tr>
<tr>
<td>kolgokoro</td>
<td>9</td>
<td>1.5049</td>
<td>0.5660</td>
</tr>
<tr>
<td>shihei</td>
<td>10</td>
<td>1.6995</td>
<td>0.6885</td>
</tr>
<tr>
<td>kyotei</td>
<td>42</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>heiwa</td>
<td>37</td>
<td>0.7131</td>
<td>0.6404</td>
</tr>
<tr>
<td>kutsushita</td>
<td>3</td>
<td>1.9903</td>
<td>0.6193</td>
</tr>
<tr>
<td>kibo</td>
<td>2</td>
<td>1.9954</td>
<td>0.5545</td>
</tr>
<tr>
<td>kitai</td>
<td>16</td>
<td>1.3278</td>
<td>0.6792</td>
</tr>
</tbody>
</table>

\[0.1434 \quad 0.6555 \quad 1.0372 \quad 1.4428 \quad 1.9418\]

\[\underline{T_1} \quad \underline{T_2} \quad \underline{T_3} \quad \underline{T_4} \quad \underline{T_5}\]

J.L.: Judgemental Limit

Table 4 replicates the findings shown in Table 3. That is, all the inductive limits are within the range of the two category boundaries \(T_2\) and \(T_3\). This range corresponds to "some relationship" in the 6-point scale. These two identical results in KOC3 and YAC3 appear to give some justification to the assumption of judgemental limit or inductive limit.

We have seen above 11 items in kowasu and 10 items in yaburu indicate one thing in common, i.e., all judgemental limits are equally definable as "some relationship". This uniformity justifies the assumption of inductive limit, but it is not certain whether the assumption of projective limit is justifiable. This assumption is examined in (3). However, analysis (2) indicates that the assumption of judgemental limit itself is acceptable, and that the range of judgemental limit falls in the interval of \(T_2\) and \(T_3\).

(3)

In this section, we will try to find out whether or not the assumption of projective limit is tenable. Analysis (3) also examines chained reasoning which has been observed
widely among $L_1$ children (see chap.3). This chained reasoning has been characterized by the shift of criteria. The assumption of chained reasoning in the present context suggests that the 'sure' item (which the subjects learned in trial 1) may not be used as the basis of translatability judgement during the later trials.

Although we hoped to obtain some asymptotic performance, neither YAC3 or KOC3 was satisfactory in this respect. Table 1 & 2 in 7.2, Appendices compare the first three trials and the last three trials. For the present analysis, the last three trials were adopted since they are slightly steadier than the first three trails. On the basis of the mean frequencies, the judgemental limits during the last three trials were estimated.

Tables 5 & 6 present the cases where the subjects appear to have based their translatability judgements upon SVs with the experimental 'sure' item. 7 out of the 10 items to be judged in kowasu and 4 out of the 9 items in yaburu appear to have been judged this way. Conversely speaking, as for the other three kowasu items and five yaburu items were not judged using 'sure' items which they learned at the beginning of this experimental session, but some other items were used as the basis of translatability judgement. Since we do not know what caused the subjects to adopt these untaught items as the bases of judgement we will describe them as 'pseudo-sure' or 'subject-generated' for the time being. Yaburu elicited more pseudo-sure items than kowasu did. This may be accounted for by the dominance effect mentioned in (1). While the concrete condition promotes learning (of the initial association between item 1 and 'break'), the abstract condition tends to cause unstable learning. This might be something to do with the greater occurrence of shift in yaburu from the experimentally imposed 'sure' to pseudo-sure items.
The present estimation of judgemental limits yielded more pseudo-sure items than is necessary. Tables 7 & 8 present these new possible association chains in kowasu and yaburu respectively.
KOWASU

(POSSIBLE (NEW) CHAINS WHICH DID NOT OCCUR IN TRIAL 1)

<table>
<thead>
<tr>
<th>Chains</th>
<th>SV</th>
<th>DD</th>
<th>J.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 → 9</td>
<td>1.3977</td>
<td>0.4737</td>
<td>1.4072</td>
</tr>
<tr>
<td>7 → 2</td>
<td>1.8594</td>
<td>0.3989</td>
<td>1.3767</td>
</tr>
<tr>
<td>11 → 2</td>
<td>0.7662</td>
<td>0.4813</td>
<td>1.3729</td>
</tr>
<tr>
<td>11 → 5</td>
<td>1.5535</td>
<td>0.5410</td>
<td>1.1712</td>
</tr>
<tr>
<td>4 → 6</td>
<td>1.6378</td>
<td>0.4025</td>
<td>1.4527</td>
</tr>
</tbody>
</table>

J.L.: Judgemental limit

Table 7

YABURU

(POSSIBLE CHAINS WHICH DID NOT OCCUR IN TRIAL 1)

<table>
<thead>
<tr>
<th>Chains</th>
<th>SV</th>
<th>DD</th>
<th>J.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 → 3</td>
<td>1.0335</td>
<td>0.6567</td>
<td>0.8825</td>
</tr>
<tr>
<td>6 → 1</td>
<td>1.4490</td>
<td>0.5786</td>
<td>0.9514</td>
</tr>
<tr>
<td>6 → 7</td>
<td>0.4835</td>
<td>0.6181</td>
<td>0.6566</td>
</tr>
<tr>
<td>10 → 1</td>
<td>1.2800</td>
<td>0.6802</td>
<td>0.9671</td>
</tr>
<tr>
<td>10 → 3</td>
<td>1.0078</td>
<td>0.6175</td>
<td>0.8658</td>
</tr>
<tr>
<td>10 → 4</td>
<td>1.0324</td>
<td>0.5771</td>
<td>0.7035</td>
</tr>
</tbody>
</table>

Table 8

Table 7 (kowasu) shows that items 4, 7 & 11 are pseudo-
sure items which the subjects appeared to have produced
during the last three trials. Table 8 (yaburu) indicates that these subject-generated sure items are items 6, 7, 9 & 10. It remains to be seen whether these pseudo-sure items are the outcome of reasoning on the part of the subjects. If we can recognize some principles in the way in which they change their bases of judgement, we will regard them as sure items which are generated by the subjects, based upon sound inference. But it is arguable whether they are really 'sure' items in the mind of the subjects, since the present hypothetical argument depends on how plausible the assumption of projective limit is.

In the above we noted that the subjects reproduced initially learned associations in seven cases of kowasu (see Table 5). The present method yielded five new chains (Table 7). We have thus two unnecessary chains. In the case of yaburu, we have four unnecessary chains. We may account for these superfluous chains two ways. (a) We assume that some items involve two "sure" items. This assumption is not contradictory to the other assumptions of paired comparison and judgemental limit. But it is not desirable to accumulate assumptions one after another, everytime we face a new problem in the data. (b) The second approach is to analyse the clear and obvious cases. There are a few obvious cases in which some associative chains which occured before disappear completely in later trials and so on. Since we are investigating analogical processes (§4.1), we will look at the influence of SV on these clear cases. Thus, in terms of SV, we compare the initial associations (which occurred in trial 1 but were replaced during trials 3-5) with the new possible associative chains. This simple comparison yielded an interesting finding; i.e., the shifts mentioned above have some underlying principles which appear to have led to their selection.

Table 9 compares the discarded associations with the new chains which replaced the initial associations. Table 9 deals with kowasu.
Old & discarded chains  New replacements

(KOWASU)

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>DD</th>
<th></th>
<th>SV</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>1.4655</td>
<td>0.4743</td>
<td>7-9</td>
<td>1.3977</td>
<td>0.4737</td>
</tr>
<tr>
<td>1-6</td>
<td>2.0232</td>
<td>0.3901</td>
<td>4-6</td>
<td>1.6378</td>
<td>0.4025</td>
</tr>
</tbody>
</table>

(YABURU)

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>DD</th>
<th></th>
<th>SV</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-9</td>
<td>1.9954</td>
<td>0.5545</td>
<td>6-9</td>
<td>1.4306</td>
<td>0.5804</td>
</tr>
<tr>
<td>7-9</td>
<td>1.1416</td>
<td>0.5521</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9

It is evident that discarded associations are those whose SVs are greater than the SVs of the replaced associations which remained as possible chains in later trials. Tables 10 & 11 compare the initial associations which remained as chains in trials 3-5 with the newly created chains.

Old & possible  New & possible

(KOWASU)

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>DD</th>
<th></th>
<th>SV</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>1.7076</td>
<td>0.3641</td>
<td>11-2</td>
<td>0.7662</td>
<td>0.4813</td>
</tr>
<tr>
<td>1-5</td>
<td>1.9446</td>
<td>0.3425</td>
<td>7-5</td>
<td>1.3149</td>
<td>0.4868</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td>11-5</td>
<td>1.5535</td>
</tr>
</tbody>
</table>

Table 10

(YABURU)

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>DD</th>
<th></th>
<th>SV</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>1.4822</td>
<td>0.5669</td>
<td>10-1</td>
<td>1.2800</td>
<td>0.6802</td>
</tr>
<tr>
<td>2-3</td>
<td>1.1662</td>
<td>0.8083</td>
<td>10-3</td>
<td>1.0078</td>
<td>0.6567</td>
</tr>
<tr>
<td>2-4</td>
<td>1.5049</td>
<td>0.5660</td>
<td>10-4</td>
<td>1.0324</td>
<td>0.5771</td>
</tr>
<tr>
<td>2-7</td>
<td>0.7131</td>
<td>0.6405</td>
<td>6-7</td>
<td>0.4835</td>
<td>0.6181</td>
</tr>
</tbody>
</table>

Table 11

362
Tables 10 & 11 also indicate that except for the chain 7 - 2, all newly created chains are semantically closer than the initial associations which are still possible candidates in trials 3-5. Since this principle of semantic closeness is observed equally in kowasu and yaburu, this may be one of the governing principles which determines the selection of inferred 'sure' item, as opposed to the learnt 'sure' item.

Assuming this is the operative principle, it is likely that the newly created chains in Tables 10 & 11 are real chains and that the initial associations were actually discarded by the subjects. Once we remove these chains, no longer regarding them as possibilities, item 7 in kowasu remains as an authentic candidate of a 'sure' item, and in yaburu items 6, 5, 8 & 10 remain as authentic. Diagrams 1 & 2 summarize the discussion so far. Arrows indicate the same association that occurred in trial 1. The dotted arrows represents the chains which were created during trials 3-5.

(...see Diagrams 1 & 2, in 7.0, Appendices).

The following routes of inference illustrate chained reasoning in a proper sense of the term.

(i) 1 → 11 → 2 (kowasu)
     \   \    \n     5 1

(ii) 6 → 7 → 3 (yaburu)
     \   \    \n     9 4

(iii) 10 → 4 (yaburu)
      \   \    \n      1

Diagram 3a chained reasoning
What we have called chained reasoning here are classified into two classes of chained association and wholistic association in L₁ studies; and the former is subdivided into combinatorial association, chain complex and associative complex (see chap.3). The following 4 cases are instances of chain complex, according to the definition offered in chap. 3.

(a) 6 → 11 → 2  (c) 6 → 9 → 4
(b) 6 → 7 → 3  (d) 6 → 7 → 9 → 4

Diagram 3

The following four cases are instances of associative complex or combinatorial association, according to the respective definitions presented in chap.3.

(a) 11 ← 2
    5

(b) 6 ← 7
    1

(c) 7 ← 3
    9

(d) 10 ← 4
    1

Diagram 4

We cannot determine whether the above cases in Diagram 4 are regarded as associative complex or combinatorial association. This is because the definitive answer to this question is obtainable, only when we know for certain what sorts of semantic features the subjects actually based their inference on. This mystery of what is inside the black-box one's brain is beyond the scope of the present research. But it is at least evident that two or three kinds of chained reasoning (which were observed among L₁ children) are detectable among the present L₂ subjects. The four kinds of associative chains are noted as reflecting the creative force of language by Waldron (1979) from the
view point of deachronic change of word-meaning (see chap.3). Labov's longitudinal researches indicate that the diachronically observed change occurs synchronically within the span of one's life. The present results seem to indicate that the basic principles governing diachronical changes can operate almost instantaneously for less than 45 minutes. This observation has some implication to the controversial issue of whether language acquisition occurs instantaneously or develops gradually in the course of time. This controversy was raised by Chomsky (1976). The present result implies that the controversy may be interpreted somewhat differently. That is, the two assumptions instantaneous or gradual, do not differ much, once we regard that the basic principles are in operation both instantaneously and gradually in the progress of time. When we look at performance data which were collected longitudinally, the result may appear to look as if learning took place either instantaneously or gradually. This impression may be the outcome of the researcher's point of view who overly obsessed with this controversy and the data behaviours on the surface. Underneath the apparenr tendency which performance data indicate, there might be cognitive principles in operation which give rise to the observable kinds of performance. But if the above controversy were a cognitively meaningful issue, the present result appears to support the instantaneous assumption. The assumption of projective limits seems to be fairly tenable, but, the assumption of judgemental limit has difficulty in supplying a plausible explanation for these enigmatic items. This is examined in (4) below.

(4)
One noticeable characteristic in common with items 7 in kowasu and items 5, 6, 8 & 10 in yaburu is that we cannot trace the route of inference back to limital 'sure' items 1 (in kowasu) & 2 (in yaburu): see Diagrams 1 & 2, 7.0 Appendices. They look as if they occurred independently during trials 35. However, by tabulating all items on the respective lists according to SVs, we notice that those items which are untraceable to the initially instructed
items are the most similar to the initial items in terms of SVs. Here, we can recognize that a principle of maximal similarity is in operation.

<table>
<thead>
<tr>
<th>Kowasu Sense distance from 1</th>
<th>Yaburu Sense distance from 2 in each cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>DD</td>
</tr>
<tr>
<td>2</td>
<td>1.7076</td>
</tr>
<tr>
<td>3</td>
<td>1.5800</td>
</tr>
<tr>
<td>4</td>
<td>1.6427</td>
</tr>
<tr>
<td>5</td>
<td>1.9446</td>
</tr>
<tr>
<td>6</td>
<td>2.0237</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>1.8322</td>
</tr>
<tr>
<td>9</td>
<td>1.4655</td>
</tr>
<tr>
<td>10</td>
<td>0.8997</td>
</tr>
<tr>
<td>11</td>
<td>1.8052</td>
</tr>
</tbody>
</table>

(The four clusters A-C' corresponds to those in the MDS configuration.)

Table 12

It is reasonable to say that item 7 (kowasu) and items 6 & 10 (yaburu) are chosen by the subjects according to the principle of maximal similarity. The principle of maximal similarity may be seen as an important part of analogical process. So far we have come across two principles which relate to this analogical process: (1) semantic closeness and (2) maximal similarity. As we see below (see also §10.2), there is another principle called "identity principle". For these reasons it appears that similarity standards which Quine (1976) regards as "innate" in reference to L₁ acquisition are important in L₂ acquisition.

Every time we rearrange the items according to SVs, we have discovered some regularity in the data. This seems to suggest that the subjects themselves organize the items on the list repeatedly according to their sense of similarity, until they arrive at their conclusion of translatable-
ty. It is likely that the judgemental process involves this iterative reorganization of items based upon the similarity principles mentioned above. (When the free recall experiment was run jointly with the translatability judgement, the same patterns of reorganization were observed (see chap.6).) The items we have examined in relation to Table 12 have the following characteristics: (1) they cannot be traced back to the initial stimuli, by the method of judgemental limit; (2) they are maximally similar to the respective initial stimuli. Once we supply some explanation regarding (1), it will follow that they are subject-generated 'sure' item as a result of successive inference during the five trials. If so, we are also entitled to say that the assumption of projective limit is plausible with the conditional statement that similarity standards comes into the judgemental process.

At this point, the hypothesis mentioned in §10.1.3 (4) comes into our discussion. The prediction that kowasu has one uniquely determined relative complement with respect to the initial stimulus, item 1 is confirmed. Furthermore, the prediction that yaburu possesses more than one relative complement with respect to the initial stimulus, item 2, is also born out. It is highly likely then that items 5, 6, 8 & 10 are relative complements on the part of the subjects. If so, the prediction based upon the assumption of modular lattices is born out also in the case of yaburu. The predictions based upon the difference between modular and distributive lattices were confirmed by the data. This suggests that item 7 (in kowasu) and items 6 & 10 (in yaburu) are the outcome of inference on the part of the subjects. In this sense, these items are authentic subject-generated 'sure' items.

Following Piaget (1953), we can assume that lattices specify cognitive operations explicitly. The above relative complement corresponds to Piaget's "groupment 2". What Piaget calls "vacariance" is especially relevant to the present discussion. Piaget's formula (A+A'=B) represents the operation of vacariance. As an illustration (see Diagram 5),

367
we assume A is the initial association of TL sign 'break' with one of the NL concepts (e.g., item 7) which the subjects learned. A' is a complement relative to A. In kowasu, for instance, A' is item 7.

A' (item 7) vicariously substitutes the initially learned association A in the later trials, and becomes a 'sure' item whose translatability the subject conceives as established. This newly created 'sure' item then comes to be used as the basis on which the translatability of other items are going to be judged.

This operation of vicariance may occur repeatedly. This iterative application will result in a hierarchical structure such as follows:

This suggests that the operation of vicariance is one of the sources which brings about differentiation. For in-
stance in Diagram 5 three items are distinguished, but in Diagram 6 seven items are differentiated. The preset argument may be related to the differentiation discussed in §9.3.2 (4). In §9.3.2 (4) we looked at the MDS configurations developmentally, and recognized the change from differentiation to unification in kowasu and from unification to differentiation in yaburu. These two terms were used to describe the state of configurations, the outcome of processing. But in the above, we have argued that differentiation stands for a on-going process which possesses some explicit underlying cognitive principles, such as vacariance.

If so, we might well speculate whether unification is also principled a process rather than a accidental phenomenon which happens incidentally. Unification is in opposition to differentiation, suggesting that unification involves negation: (N) in Piaget's INRC. We have also seen in §9.3.2 (4) that unification involves generalizing a particular feature, by simplifying minor differences. This simplification relates to Piage's negation. Piaget's INRC has negators called N; and R stands for reciprocal operation and C, correlative. Apart from I (identity), N, R and C bring about some aspect of negations:

In §3.4 we have seen that word-meaning can be represented by $\sum b_i$ and $\prod b_i$ in conjunction with INRC. Unification seen in relation to negation (N or C*R) would involve lexical reorganization (see §3.4). It appears that the joint and complex functions of N, R and C will contribute to the formation of unification. This is a complex issue and one of further topics for research. But the brief reflection suggests that unification involves
cognitive principles which can be formulated explicitly. In this sense, unification is a process. Since unification can be explained by the complex interaction of INRC, we may say that unification is a cognitively principled behaviour.

It is now clear that the subject-generated 'sure' items are the outcome of vacariance and that they are characterized as relative complements of the initially instructed 'sure' items. We have also argued in §10.1.3 (4) and above that the assumption of relative complement makes sense in the context of translatability judgement. For this reason, the theory of lattices is generally useful in dealing with the process of translatability judgement.

So far we have explained a large portion of pseudo-sure items and seen that they were the result of principled behaviours. Except for items 5 & 8 in yaburu, the assumption of projective limit has stayed intact. We will examine below whether or not items 5 & 8 are the result of principled cognitive activities.

The relationship between structures of lattices and the MDS configurations is useful here. It appears that some examination of this relationship helps us to characterize items 5, 7 & 8. I have previously referred to Shepard's and Arabie & Boorman's view that the MDS configurations are closely connected with the structures of lattices (see above). The MDS configuration of YAC3 presents the shape of an oblong square. This shape is a slightly deformed modular lattice of the following kind:

![Diagram 8](image)
(In the MDS configuration, item 10 is overlayed with items 9, 4 & 1, and item 7, with items 6, 3 & 2.)

According to Matsumoto, a, b, c, d & e are all equivalent (identical) to each other in the structure. We have so far two similarity-oriented principles (of semantic closenes and maximum similarity). Since the MDS map has been discussed in view of lattices, we might go even further and include Matsumoto's proof of equivalence as another governing principle here. We will call this principle of identity (Matsumoto uses the term 'equivalence' in the sense used in logic, but in terms of SV it has magnitude and so it is more appropriate to call it 'identical'.) This assumption of an identity principle predicts that items 5, 8, 7 & 10 are conceived of by the subjects as 'sure' items. This prediction fits in with the data. The subjects regarded them as 'sure' items (see Table 12).

The above principle of identity requires two comments. (1) It represents a highly economical processing. (2) It relates to unification mentioned in §9.3.2 (4). As for the former, we may note that all four points can be picked out immediately by the subjects, when they adopt this principle, because the four stimuli (which are represented as the four corners in the 2-dimensional space) lattices. In this sense, the principle of identity offers a highly economical method of data-processing. The present discussion also suggests that unification is not a state, but a process. The MDS configuration reflects the outcome of unification process. With regard to the second point, we can indicate the following. The modular lattices relates to the kind of unification shown by YAC3. It is also true to say that distributive lattices show regular structures (see Takeuti (1975) & Matsumoto (1980)). As long as the subjects are adopting the logical principles of lattices, the process of unification relates to both distributive and modular lattices.

In the above we have ascerained that both differentiation and unification represent processes. It may sound as if
two processes are different. But from the view point of lattices, they are dual. This means that when differentiation occurs, unification occurs. For instance,

(i) \( x \cup (y \cup z) = (x \cup y) \cup z \)

Then,

(ii) \( x \land (y \land z) = (x \land y) \land z \) holds true.

The left side of eq.(i) corresponds to \( A + A' = B \) in Piaget's formula. We regarded this as one of the sources of differentiation. When this differentiation happens, the duality principle indicates that eq.(ii) holds true. Since we regard (i) as a source of differentiation, (ii) which is represented by (the opposite of \( U \) in (i)) is a source of unification. This suggests that unification is the outcome of differentiation and vice versa. This point will be illustrated by Experiment 2 below.

It may be recalled that we have postulated iterative reorganization in this section. The experiments in chap.6 confirmed this hypothesis. We have also argued that this iterative reorganization would be carried out on the basis of such similarity-closeness and maximal similarity. We now also include the principle of identity. Then When the subject reorganizes the NL items in order to arrive at his conclusion of translatability, he would rely on all three of these similarity-based principles. Thus, it is highly likely that he conceives of items 5, 8, 7 & 10 as 'sure' items and uses them as the bases of his translatability judgements. This accounts for the fact that items 5 & 8 cannot be traced to any other 'sure' item. The subject does not need any other 'sure' item in order to deduce his conclusion, because the items themselves are already 'sure' items in his mind.

Analysis (3) included item 7 as part of the chained reasoning ( 6 \( \rightarrow \) 7 \( \rightarrow \) 3 or 6 \( \rightarrow \) 7 \( \rightarrow \) 9 ). However, according to the above principle of identity, item 7 is a 'sure' item. For this reason, Diagrams 2, 3 & 4 need to be modified, but only the modified version of Diagram 2 (YAC3)
is presented here.

![Diagram 9](image)

The most noticeable feature in the above reconstruction of inference routes is that the initial instruction (yaburu) is no longer conceived of by the subjects as a 'sure' item. This may be another reflection of the dominance effect examined in (1).

The judgemental process involves differentiation. This process of differentiation is seen in both verbs. Diagram 1 (kowasu) shows that a set of eleven NL items is divided into two with the two 'sure' items 1 and 7 at the foci of the two clusters. Diagram 5 (yaburu) involves six isolated groups. The connections among these groups are not entirely clear. In §9.3.2 (4) we have observed the movement from differentiation to unification in kowasu, and the reverse movement from unification to differentiation in yaburu. The present discussion suggests that these terminologies can also describe processes as well as the products of processes, as we have indicated previously.

KOC3 was more straightforward to analyse, probably because the NL interlexical structure (shown in the MOS result) is more well-structured than YAC3. KOC3 produced only one subject-generated 'sure' item (item 7) which reflects that the underlying structure is that of distributive lattice, since the prediction made in §10.1.3 (4) is born
out. The choice of item 7 appeared to be governed by the principle of maximal similarity and confirms that a distributive lattice involves only one unique complement in relation to the initial stimulus. In the case of YAC3, there were five subject-generated 'sure' item. Of these five, items 6 & 10 appeared to be determined by the principle of maximal similarity. Since modular lattices have more than one relative complement, this theoretical prediction was born out by the data. The characteristic feature of modular lattices also accounts for the four subject-generated 'sure' items (5, 8, 7 & 10). The occurrence of these items (5, 8, 7 & 10) are explained by the principle of identity which we derived from Matsumoto's proof on modular lattices. In this way, we have managed to clarify all pseudo-sure items in the data. As we have seen in §9.2 (see below), the subjects appear to have adopted categorical processing which includes all other processings as its integral parts, and to have manipulated them as the occasion arose, to cope with the various aspects of the experimental task. For this reason, the analytical framework must be flexible enough to accommodate both modular and distributive lattices.

The analysis showed that it was necessary to postulate three kinds of similarity: (1) semantic closeness; (2) maximal similarity; and (3) identity. The second of these was useful in distinguishing two kinds of 'sure' item: one is a relative complement in distributive lattices and the other, the relative complement in modular lattices. The third principle of identity (derived from modular lattices) explained the four 'sure' items in vaburu. The meaning of identity used in modular lattices is very different from the meaning of equality in our arithmetics which follows distributive lattices (see below). SIM is distributive in this sense and deals with the three principles in its own way, but we cannot expect SIM to reveal the structure of modular lattices transparently. Yet, some of the subject-generated 'sure' items requires the notion of identity which originates from the structure of modular lattices. We will discuss this point further.
below, but first of all we will summarize our argument concerning inductive and projective limits.

The above discussion confirms three points.

(1) The assumption of projective limit is plausible, because all basic guidelines are provided for by the projective limits estimated.

(2) But the naive assumption of projective limit also is not tenable. We need some auxiliary principles to reveal the analogical processes. The principles shown above are (a) semantic closeness, (b) maximal similarity and (c) identity. When we incorporate these principles into the assumption of projective limit, the resultant analysis clarifies the data behaviour.

(3) As to the effects of processing upon the judgemental process, we have considered the effects in relation to the two equations presented in chap.9. Namely, the version of dominant process shown in YAC relates to modular lattices and K03, distributive lattices. This difference in lattice structures clarified the analysis to a large extent in that (a) it can account for all pseudo-sure items, and (b) that it bridges the gap between processings and operative principles mentioned in point (2). The effect of processing upon the judgemental process can not be presented in a straightforward manner. The effects of processing become clear only when we find some common ground not only for the analogical process but also for the underlying mechanisms of processing. We found this common ground in the lattices or Boolean algebras.

We have induced the principle of semantic closeness on the basis of Tables 10 & 11. This particular principle of semantic closeness falls in the range of T₂ and T₃, although the range is narrower than that of judgemental limits. We have also induced the principle of maximal similarity, based upon Table 12. This particular principle
is below $T_2$, according to Table 12. Then, the third principle was postulated in relation to a modular lattice whose structure is diamond-shaped. In terms of SVs, this principle is less than $T_1$. In this sense, we have the following inclusion relationship among these principles:

```
identity maximal similarity semantic closeness

T_1 ................ T_2 .... T_2 - T_3 ...........
```

[Diagram 10]

I analysed the data, starting with the principle of semantic closeness, and making use of judgemental limits and SVs. By working through the principles which involves the closer similarity and the closest similarity, one by one, we have managed to explain our data. So, our investigation started with the right end of the arrow and ended up with the identity principle, the inner most subset of similarity set (see the top shematic diagram). It is reasonable to suppose that the subjects should take the opposite route, starting with the identity principle and ending in the semantic closeness principle. This is because it is conceptually easiest to intuit identity rather than the semantically remoter principles of maximal similarity and semantic closeness. This hypothetical argument is
reasonable in view of our empirical study of free recall in relation to the translatability judgement test (see chap.6). The principle of semantic closeness relates to large chunking, the principle of maximal similarity, to smaller chunking and paired comparison, and the principle of identity, the smallest chunking and paired comparison. These three kinds of chunking and 2-item clusters were obtained from the free recall data. In this sense, the above three principles were supported by the experiments in chap.6. We could not however conclude which chunking comes first, although we speculated the order of occurrences from large chunking to the smallest chunking consisting of two items (see strategies 1 & 2 in chap.6). When categorical processing is at issue, all these chunking can occur simultaneously, since this processing is characterized by parallel processing (see §9.3.2(2)). Further, chap.9 concluded that all the five groups of subjects adopted categorical processing. This suggests that Diagram 10 is misleading in that the order of occurrences (of both chunkings and principles) is linearly presented. When these chunkings and principles operate in parallel, the logical framework we need is a topological Boolean algebra (which has been investigated by the present experimenter but has not been included in this thesis).

For the sake of convenience for our discussion in §10.2, it is appropriate to summarize the relationships between eq. 5 & 6 in chap.9 and the five kinds of processing:

Modular lattices

<table>
<thead>
<tr>
<th>eq. 5</th>
<th>eq. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(separable (finite))</td>
<td>(integral)</td>
</tr>
<tr>
<td>(asymmetric separable</td>
<td></td>
</tr>
<tr>
<td>(finite/one compelling dimension))</td>
<td></td>
</tr>
<tr>
<td>(dominant (infinite/one compelling dimension))</td>
<td></td>
</tr>
</tbody>
</table>

Categorical

Diagram 11
When eq. 5 & 6 are considered in relation to the distinction of modular and distributive lattices, we get Diagram 12 shown in §10.2.1 (requires eq. 5 & 6). Reflecting performance factors in chap.9, KOU2 was highly likely and YAU2 was likely to have adopted separable processing; and both YAC3 and YAK1 adopted dominant processing. It should be born in mind that all subjects essentially adopted categorical processing. The above specific processings are due to the immanent performance factors which were largely caused by the experimental situations. For this reason, we must be prepared to face the possibility of parallel processing (inherent in categorical processing). That is, our method of analysis in §10.2 should make provisions to incorporate the four kinds of processing which the categorical processing subsumes.

As Diagram 11 (see also §9.3.2(2)) indicates, the only difference between separable and dominant processings is whether the given object under investigation is finite with respect to any number of dimensions or infinite with respect to one dimension. The only difference between dominant and asymmetric separable processing is whether the given object under investigation involves an infinite operation or a finite operation. Eq.5 can be finite or infinite (see §9.2). For this reason, it is more effective to reduce the three types of processing to eq.5 which corresponds to modular lattices (see Diagram 12 in §10.2). Likewise, the various computational operations included in integral processing are reducable to the pair of eq.5 and eq.6 which corresponds to distributive lattices. Thus, it is clear why categorical processing which is represented by the pair of eq.5 & 6 includes the four kinds of processing simultaneously.

At the same time, the distinction of modular lattices and distributive lattices is sufficient for our purpose without making finer distinctions of dominant, separable and asymmetric separable processings. If a given object has the structure of distributive lattices, the object implies that it has a structure of modular lattices as well.
(according to Matsumoto's theorem 1.3.3). The structure of modular lattices (which is relevant for the present research) involves "identity", as we have seen in this section (point (4)). The principle of identity, in this case, is evidence for the three kinds of processing which are subsumed under the structure of modular lattices. On the other hand, the integral processing which has been connected with the Eucliden metrics is eligible for such arithmetic operations as +, -, \( \times \), \( \div \) and more advanced computation of differentiation, integration, and so on. (One of the reasons why Shepard says the Euclid metrics is robust appears to be that the Euclid metrics involves the pair of eq.5 & 6 and that it is suitable for the usual mathematical operations listed above.) Now, SIM requires most of these usual mathematical operations. This is the limitation of SIM, since SIM can pin down the principle of identity, only when a particular SV is less than \( T_1 \) (which corresponds to the nominal range of identity in the 6-point scale we have adopted). This is not satisfactory.

(But, the frequency scores of translatability acceptance in the sense of counting assumes distributive lattices; without this assumption, we cannot count scores nor compare the frequency of scores. In this sense, both SIM and the counting of acceptance scores implicitly assumes distributive lattices.)

Our estimation of judgemental limits also assumes SIM and therefore assumes distributive lattices. Furthermore, SIM computes judgemental limits on the basis of acceptance scores of translatability which implicitly assume distributive lattices. This suggests that in order to obtain better estimation of the principle of identity, we need to take into account not only \( T_1 \), but also arithmetic approximation (rounding): i.e., the difference of 1 or 2 in acceptance scores may be ignored. Also an equal number of acceptance scores may be taken as an indication that the principle of identity is in operation. These operationally defined principle of identity may accommodate the
conceptual (cognitive) identity expressed by modular lattices (see Diagram 8 of this chapter).

Thus, since the present method is based upon distributive lattices we will make the following provisions to incorporate the principle of identity which is derived from modular lattices. The principle of identity for modular lattices is defined as

(1) the judgemental limit being less than $T_1$ (the boundary specifying "identical");

(2) the equal number of acceptance scores being taken as an indication of the principle (coded as $=$ below);

(3) involving the following arithmetic roundings:

   (i) the difference by one (coded as $\pm$ below)

   (ii) the difference by two (coded as $\pm 2$ below).

In what follows we assume the above operational approximation gives a fair approximation of "identity" derived from modular lattices.

§ 10.2 experiment 2 (developmental trends)

Experiment 2 deals with kowasu and yaburu among four groups of subjects: first year high school students (coded as KOK1 and YAK1); second year high school students (coded as KOK2 and YAK2); first year university students (coded as KOU1 and YAU1); second year university students (coded as KOU2 and YAU2). There were 45 students in each group. These subjects all took part in the SIM experiments (chap. 5) and the experiments in chap. 9. Thus, we have four sets of SVs, DDs and MDS results for each verb. Since the method of administration of the translatability judgement test was the same as in KOC3 and YAC3 except for one instruction, it is not presented here. They were not provided with any experimental 'sure' item.
§10.2.1 gives the aim and hypotheses of this experiment. §10.2.2 discusses the results.

§ 10.2.1 aim and hypotheses

Experiment 2 is cross-sectional. For this reason, the main interest of this experiment is to look at the developmental trend among the four groups of subjects. The development is studied after treating the data three ways (1)-(3) shown below. We regard some assumptions which were confirmed by Experiment 1 as tenable. On the basis of these assumptions confirmed by Experiment 1, we set up further hypotheses for Experiment 2.

Data treatment 1 - 3

(1) Since the method of estimating inductive and projective limits was confirmed as adequate by Experiment 1, we will adopt the same method and make use of SVs and DDs to estimate these limits. Experiment 1 also confirmed that the range of all judgemental limits was in the interval of T2 and T3. This range notionally translates into 'some relationship' in the 6-point scale which the subjects were presented with for the SIM experiment (chap.5). Based on this finding, we compute a set of judgemental limits separately for each group. This enables us to reconstruct the routes of inference in the same manner as we have done in the case of KOC3 and YAC3 (see Diagrams 1 & 2).

(2) As we have seen in §10.1.4, one weakness of SIM is that it is not adequate in estimating identity in the sense of modular lattices. In order to make up for this weakness, we proposed the operational definitions of identity so that some indication of equality in the acceptance scores would be included in the category of 'identity'. Consequently, we regard as identical not only those whose judgemental limits are below T1 ("identical" in the 6-point scale), but also the same number of acceptance scores and those whose acceptance scores differed by only one or two. Thus, we have three
indications of 'identity' and they are coded as \( T_1 \) (or \( \text{ident.} \)), \( = \), \( \ell \), and \( \ell^2 \) in the reconstructed maps of inference routes.

This provision was necessary, because the principle of identity derived from modular lattices represents a highly economical processing. As I suggested in §10.1.4 (Diagram 8), the principle of identity appears to have enabled the subjects to intuit, at once, the semantic identity among all four stimuli (each of which corresponds to one of the four corners of the square in the diagram): see Diagram 8. Thus, from the subject's view point, the principle of identity is important. But this provision for the identity principle is also significant from the experimenter's viewpoint, since it encompasses the situation in which eq. 5 operates on its own. As Diagram 11 in the previous section indicates, the independent operation of eq. 5 captures the essentials of the dominant and asymmetric separable processings. This suggests another important data treatment. Diagram 11 shows that one can tabulate the data, according to the dichotomy of modular vs distributive lattices:

<table>
<thead>
<tr>
<th>Modular lattices</th>
<th>Distributive lattices</th>
</tr>
</thead>
<tbody>
<tr>
<td>eq. 5</td>
<td>eq. 5 &amp; 6</td>
</tr>
<tr>
<td>dominant</td>
<td>integral</td>
</tr>
<tr>
<td>asymmetric</td>
<td>separable</td>
</tr>
<tr>
<td>separable</td>
<td></td>
</tr>
</tbody>
</table>

Categorical

\[ =, \ell, \ell^2 \text{ and } T_1 \]

Identity

\[ T_1 \rightarrow T_2 \text{ and } T_2 \rightarrow T_3 \]

Maximal & Some

(Semantic closeness)

Diagram 12

As we have seen in the previous section, we can tabulate the data according to the division of \( =, \ell, \ell^2 \text{ and } T_1 \) on one hand and those whose judgemental limits fall in the range of \( T_1 \text{ to } T_2 \) and the range of \( T_2 \text{ to } T_3 \) on the other. Further, this division indicated by the vertical line in Diagram 12 corresponds to the distinction between the
principle of identity and the other two principles. In this sense, this data-treatment (2) is useful in studying two interesting relationships: modular vs distributive lattices and the principle of identity vs the other two principles of maximal similarity and semantic closeness.

The basic procedure of data-treatment (2) consists of three steps. (a) Translatability acceptance scores will tell us which satisfy =, =, and 2. (b) The estimates of judgemental limits will show which items will satisfy T1. (c) Similarly, the estimates of judgemental limits will show which items will satisfy the principle of maximal similarity (T1 - T2) and the principle of semantic closeness (T2 - T3).

(3) We are interested in the "U-shaped behavioural growth." This U-shaped curve is known to be related Piaget's framework (see Strauss, 1982). Although we now know that the judgemental process relates to the modular and distributive lattices which Piaget adopts in his eight groupments, we have not managed to connect lattices to the U-shaped curve in this thesis. Since no explicit formalization was attempted here, we will analyse the data with respect to the U-shaped curve in a naive way.

Strauss writes: "Calling the phenomenon U-shaped is arbitrary ... in and of itself the shape of a curve is not particularly interesting": (1982:1). This is because the shape of a graph depends on how we plot the data. If we use a coarse unit, it will not be possible to draw a curve; the plot will instead resemble a scattergram or a broken bar chart. Since we do not have any provision to adopt a non-linear equation, we will merely plot the acceptance scores on a chart.

The apparent feature of this U-shape curve is: (a) a certain behaviour appears; (b) it disappears; (c) it appears again (see Strauss (1982)). This appearance-disappearance - appearance pattern has been interpreted variously as involving "representational reorganization". In this sense, it is a very interesting topic to pursue, but in this thesis I am only concerned with whether the "U-shaped" pattern is observable or not.
Hypothesis 1: the judgemental reflection of differentiation and unification would be ambivalent for the following reasons. In §9.3.2.(4), we have seen that the developmental examination of the MDS configurations (which reflect the 'different' use of processing to accomplish the experimental task according to the five age-groups) involved the movement from differentiation to unification in the case of kowasu, and that in yaburu the shift was reversed. We have seen in the latter the movement from unification to differentiation, although YAU2 showed some indication of unification. We have argued in §10.1.4.(4) that one of the sources of differentiation is Piaget's 'vicariance' which is based upon his principles of logic. This suggested that differentiation is a process rather than a state. If so, this process of differentiation may be seen in the manner in which the subjects infer translatability.

On the other hand, YAC3 illustrated the state of unification according to the result obtained in §9.2.3 (4). As in KOC3, the MDS configuration is the result of processing and the particular processing YAC3 appear to have adopted was dominant processing which is represented by eq. 5. This strongly suggested to us that the configuration represents the structure of a modular lattice. Then, the state of unification is the outcome of modular operation. In this sense, the state of unification implies that the process of unification has occurred previously. For this reason, it is reasonable for us to expect that both unification and differentiation would be reflected in the judgemental process. This examination is made possible by looking at the eight maps in which the inferences routes are reproduced by the data-treatment (1) above.

If the underlying operations the subjects adopt are actually based upon distributive lattices (categorical processing), it implies two things. (1) Because of the duality principle, the examination of either unification or differentiation will suffice for the evidence of the other. (2) More important, as indicated in §10.1.4,
according to the duality principle the presence of differentiation entails the dual operation of unification and vice versa. Then, it suggests that the two opposite processes would mutually cancel out the effects and the result would not show any clear evidence of one or the other.

Hypothesis 2: vaburu would yield heavier reliance on modular lattices (MOD) and kowasu, on distributive lattices (DIS). Similarly, vaburu will show heavier reliance upon the identity principle than kowasu.

This hypothesis relates to data-treatment (2). This treatment of data is convenient in indicating the following difference: i.e., how often the subjects have used processings related to either modular or distributive lattices (see Diagram 12 above). Chap. 9 showed that all groups essentially adopted categorical processing, but in terms of performance, specific processings were distinguished. We have seen that vaburu related to dominant processing, and that kowasu, in comparison, to separable processing on the whole. This general tendency can predict that vaburu would yield heavier reliance on modular lattices than kowasu would. Our prediction can be more specific in the cases which yielded high resemblances to the unit circles: KOU2 represents (asymmetric) separable processing in this respect; YAK1 and YAU2, dominant processing. In terms of the stress values, YAK2 showed the ideal case of categorical processing in the two-dimensional space. In this sense, YAK2 would show no significance difference in terms of the reliance upon one or the other kind of lattices. However, all groups essentially adopted categorical processing; only in terms of performance factors a specific processing was distinguished. Furthermore, the judgemental process may not directly reveal performance reflections of processings, since the judgement itself is another process. In this sense, the present predictions are chancy. But, if the judgemental process really reflects competence in the way we have defined in this research, we can argue that all groups would yield no significant difference in terms of
the amount of reliance upon the two kinds of lattices.

The tabulation of data above accounts for the distribution of the three principles of similarity (identity, maximal and semantic closeness). As Diagram 12 indicates, the identity principle corresponds to eq. 5 and modular lattices and the other two principles, to both eq. 5 & 6 and distributive lattices. Thus, we can present the same prediction as above with respect to yaburu and kowasu. That is, yaburu would yield the greater reliance upon the identity principle than kowasu would.

However, as we have seen in §10.1.4 (& to some extent in §6.3.3), the three principles operate in parallel. The parallel operation implies that the interplay among the three principles is relative. This suggests that we cannot make any more specific prediction. In an important sense, there is no explicit criterion which determines the most predominant principle, because they are all operating in parallel.

We have argued that the principle of identity facilitates speedy processing. For this reason, it offers an economical strategy on the part of the subjects. But the amount of reliance on this principle essentially depends upon how the subjects approached the translatability experiment. It is safe for us to concentrate on the general tendency that yaburu is more closely related with the modular lattices and the identity principle.

Hypothesis 3: the U-shape behavioural curve would predict up and down patterns in accordance with the present treatment of translatability acceptance scores.

Data-treatment (3) produces a set of four bar charts for kowasu and another set, for yaburu. When the acceptance scores (i.e., the number of subjects who regard each NL item as translatable) among many subjects are plotted along the coarse unit of one-year interval, the appearance-disappearance-appearance patterns would show up as zigzag patterns for the following reason.

The categorical processing is based upon distributive lattices. For this reason, Piaget's INRC is compatible with the logical basis of categorical processing which
the subjects have adopted at the time of making their similarity judgements (see chap. 9). The translatability judgement assumes that the process will involve assimilation of the NL and accommodation to the TL. Part of the accommodation would involve negation (N) of NL or the composition of C and R in Piaget's framework. This suggests that the translatability acceptance based upon the NL will be negated, as Diagram 13a (where we assume there are two semantic features in an NL word) shows. When this is again negated at the next stage (see Diagram 13b) to maintain some "equilibrium" in cognition according to Piaget, the same item will be accepted at the third stage (see Diagram 13b). The process here is assimilation, since the TL accommodation which occurred at the previous stage is negated now. Assuming that this successive process of accommodation and assimilation is recurrent, the resultant pattern of acceptance scores will show up in the shape of ups and downs.

- **Diagram 13a**
  - 1st stage: Appearance
  - 2nd stage: Disappearance

- **Diagram 13b**
  - 3rd stage: Appearance

**Hypothesis 4:** in the case of *yaburu* we will predict concretization which relates to the dominance effect (chap. 4 & 6 and §10.1.4) for the following reason.

We have seen that the subjects tend to learn the NL concept of *yaburu* associated with the abstract use of 'break'. The feature 'abstract' is in opposition to the feature 'concrete'. Thus, if we represent the latter as p, the former is represented as not p (¬p). According to Piaget's INRC, we can represent the process of concretization as the accommodation to the TL which triggers the negation operation, yielding not p as follows:
Assuming that \( \text{INCR} \) operates in the learner's mind, (which is quite likely, since the categorical processing he adopts is compatible with \( \text{INCR} \)), the concretization is the natural outcome of this cognitive operation.

Conversely speaking, we may postulate the opposite of concretization, i.e., to conceive an intrinsically concrete item as abstract. This is in theory possible, according to the above INRC. This time, \( p \) stands for the concrete feature. Then, 'not \( p \)' will be formed by \( N \) or the composite function of \( C \) and \( R \).

On the basis of the reconstructed maps, we can count the number of 'sure' items which are used as the bases of the subjects' translatability judgements. By examining whether the 'sure' items turn out to be concrete items, in spite of the initial input of abstract items, we can show the effect of concretization in this way.

§ 10.2.2 results & discussion

Results 1 - 4 are discussed in correspondence to hypotheses 1 - 4. Figures 1-16 (in 7.0, Appendices) represent the reconstructed maps of inference routes involving \text{KOWASU}. Figures 17-32 (in 7.0, Appendices) are those dealing with \text{VABURU}. In the former case, figures above the arrows indicate the respective judgemental limits. In the latter, the estimates of the judgemental limits are given on the right side of each Figure. Those cases (belonging to identity principle) which are picked out from the acceptance scores are given at the left hand corner of each Figure. All judgemental limits are estimated on the basis of data obtained in each trial.

Result 1: It may be noted first of all that the actual values of judgemental limits differ slightly in each trial.
But all limits conform to the expected ranges of $T_2-T_3$ (some relationship). This indicates a clear tendency that the inductive limits which the subjects induced in trial 1 served as projective limits in later trials 2 & 3. This tendency may not be as obvious as it should be. As we have seen, the judgemental process involves reorganization of NL items based upon the three principles of similarity, causing the different principles to appear in each trial. For this reason, in some cases it looks as if the subjects change the standard of judgemental limits each time. Chain 6--5 of KOK1 illustrates the point.

(............see Fig. 3-5, 7.0, Appendices)

In trial 1, the judgemental limit of this chain falls in the normal range of $T_2-T_3$, which is repeated in trial 2. (This illustrates that the inductive limit which the subjects have induced in trial 1 is used projectively in trial 2). But in trial 3, the subjects use the identity principle. This change of policy on the part of the subjects does not mean that they have randomly shifted their strategy. Looking at trials 1-3 sequentially, we notice that two clusters develop more distinctly. In trials 1 & 2, most items are connected by some inference routes or other. But in trial 3, items 2, 5, 6 & 3 are grouped into one and the remainder forms the other cluster. Although the items belonging to one cluster are linked by some inference routes, the routes do not go outside their proper domain. This growth of clear clusters illustrates differentiation. But, at the same time, we notice that the subjects have adopted the maximal similarity (only once in trial 1) and identity principle several times in each trial. It is highly likely that by using all the three principles, the subjects reorganize the NL items according to these similarity standards (see chap. 6 where we noted the recurrent reorganization of NL items which was reflected in various types of subjective organizations in the free recall results.) We have also indicated in the previous section that this iterative reorganization is likely to occur in the judgemental process, and that all the three principles would operate in parallel, when the subjects adopt categorical processing.
The more groups or clusters there are in the reconstructed inference maps, the greater amount of differentiation the judgemental process have caused to occur. For this reason, we count the number of clusters produced in the reconstructed maps. The method of counting clusters is as follows. The 'sure' items are at the origin of various chains: e.g., items 6 & 1 in KOK1. By counting the number of 'sure' items, we can make out the amount of differentiation.

According to the maps in Appendices, only one groups in kowasu indicates that the greater amount of differentiation has occurred as the trials proceeded: KOK2 begins with two clusters and ends in four clusters. The other groups are all equivocal. Numerals in brackets are the 'sure' items formed.

<table>
<thead>
<tr>
<th></th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOK1</td>
<td>2 (1 &amp; 6)</td>
<td>2 (1 &amp; 6)</td>
<td>2 (1 &amp; 6)</td>
</tr>
<tr>
<td>KOK2</td>
<td>2 (1 &amp; 6)</td>
<td>4(1,6,8&amp;9)</td>
<td>4(1,2,6&amp;9)</td>
</tr>
<tr>
<td>KOU1</td>
<td>4 (1,3,4&amp;7)</td>
<td>4(1,3,5&amp;7)</td>
<td>4(1,3,5&amp;7)</td>
</tr>
<tr>
<td>KOU2</td>
<td>3 (1,3&amp;9)</td>
<td>2 (1 &amp; 2)</td>
<td>2 (1 &amp; 2)</td>
</tr>
</tbody>
</table>

Table 13

Likewise, in yaburu, the process of differentiation is highly obscure. The following table presents the number of clusters formed in each trial.

<table>
<thead>
<tr>
<th></th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAK1</td>
<td>2(1 &amp; 7)</td>
<td>5(1,3,6,8&amp;11)</td>
<td>4(1,3,6&amp;7)</td>
</tr>
<tr>
<td>YAK2</td>
<td>3(1, 4&amp;7)</td>
<td>3(1,4&amp;7)</td>
<td>3(1, 6&amp;4)</td>
</tr>
<tr>
<td>YAU1</td>
<td>5(1,4,6,7&amp;10)</td>
<td>4(1,6,7&amp;10)</td>
<td>4(1,4,6&amp;10)</td>
</tr>
<tr>
<td>YAU2</td>
<td>4(1,4,6&amp;7)</td>
<td>3(1,6&amp;7)</td>
<td>4(1,4,6&amp;7)</td>
</tr>
</tbody>
</table>

Table 14

These obscure results confirm our hypothesis 1. The results support the view that the two simultaneous processes of unification and differentiation cancel out the clear indication of one or the other process. This suggests
that the subjects subcategorize items, retaining the memory of which superordinate category each sub-category belongs to. This relationship between subcategorization and unification relates to the duality principle inherent in lattices. In the following pairs of equations, the duality principle shows that when one of the pair holds true, the other holds at the same time.

\[(a) \ x \cap (x \cup y) = x \]
\[x \cup (x \cap y) = x \]
\[(b) \ x \cap (y \cup z) = (x \cap y) \cup (x \cap z) \]
\[x \cup (y \cap z) = (x \cup y) \cap (x \cup z) \]

Since the operations '\(\cap\)' and '\(\cup\)' occur in parallel, the differentiation and unification is likely to occur simultaneously.

\[
\begin{array}{c}
A \\
/ \\
B \\
/ \\
C \\
/ \\
(B \cap C) \\
/ \\
B \cup C
\end{array}
\]

Result 2: Tables 1-4, in 7.0, Appendices present the results concerning the distribution of MODs and DISs obtained in each group of subjects. We predicted the general tendency that yaburu would show the greater amount of reliance on MOD and kowasu, on DIS. Table 13 compares the total number of MOD summed over the four groups with that of DIS with respect to kowasu and yaburu.

<table>
<thead>
<tr>
<th></th>
<th>kowasu</th>
<th>yaburu</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD</td>
<td>82</td>
<td>116</td>
</tr>
<tr>
<td>DIS</td>
<td>75</td>
<td>58</td>
</tr>
</tbody>
</table>

\[\chi^2 = 6.5460\]

Table 15

The result supports the prediction. The subjects adopted MOD more frequently than DIS, in dealing with yaburu. The difference between yaburu and kowasu is statistically significant. This is the over-all result across the four
groups of subjects.

But the four groups differ slightly, in terms of the predominance of MOD over DIS with respect to the difference of the two verbs. We will examine this developmental difference below and discuss the other predictions mentioned in Hypothesis 1. The following four tables compare yaburu with kowasu according to the four age groups.

<table>
<thead>
<tr>
<th>KOK1</th>
<th>YAK1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD</td>
<td>25</td>
</tr>
<tr>
<td>DIS</td>
<td>22</td>
</tr>
</tbody>
</table>

\( \chi^2 = 0.1103 \) (N.S.)

Table 16 1st year high school
(............. see Table 1, 7.0, Appendices)

<table>
<thead>
<tr>
<th>KOK2</th>
<th>YAK2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD</td>
<td>28</td>
</tr>
<tr>
<td>DIS</td>
<td>15</td>
</tr>
</tbody>
</table>

\( \chi^2 = 1.0023 \) (N.S.)

Table 17 2nd year high school
(............. see Table 1, 7.0, Appendices)

<table>
<thead>
<tr>
<th>KOU1</th>
<th>YAU1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD</td>
<td>18</td>
</tr>
<tr>
<td>DIS</td>
<td>22</td>
</tr>
</tbody>
</table>

\( \chi^2 = 4.2052 \) (N.S.)

Table 18 1st year university
(............. see Table 3, 7.0, Appendices)

<table>
<thead>
<tr>
<th>KOU2</th>
<th>YAU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD</td>
<td>11</td>
</tr>
<tr>
<td>DIS</td>
<td>16</td>
</tr>
</tbody>
</table>

\( \chi^2 = 8.9766 \)

Table 19 2nd year university
(............. see Table 4, 7.0, Appendices)

In terms of the use of MOD and DIS, the first year high school students showed no significant difference in dealing with kowasu and yaburu. The same result was obtained in KOK2 and YAK2. But, both first year and second year univer-
University students used MOD more frequently with yaburu than with kowasu. These two groups yielded significant differences in this respect. No significant difference in terms of the use of MOD and DIS among the two younger groups contrasts with the frequent use of MOD among the two older groups. This contrast relates to one of the major hypotheses mentioned in Hypothesis 1. We have predicted that if the subjects revealed all the relevant linguistic knowledge by adopting the categorical processing, there would be no significant differences in the amount of reliance on MOD and DIS, since the categorical processing requires both MOD and DIS and elicits all linguistic knowledge relevant to the task in question. According to this prediction, we could say that the younger groups (of YAK1, KOK1, YAK2 and KOK2) tapped on our definition of competence (see chap. 3 and 9, the all vs some distinction of lexical competence and performance). In the light of this distinction, the older groups of KOU1, YAU1, KOU2 and YAU2 revealed 'performance' rather than 'competence', since they relied on MOD more than they did on DIS. We can explain the difference between the high school students and the university students by the developmental level of competence. While the subjects in the younger groups may not have a fully developed competence, the older university students should have more developed competence in which one can expect that there would be the greater number of semantic features.

The experimental task might not have elicited all features stored as the older subjects' competence. If their judgements were based upon some features, it accords with the present view of performance. Since the dominant processing provides them with a speedy form of processing (see § 10.1.4), it is likely that the greater number of MOD was obtained among the older subjects. On the other hand, the lexical competence among the younger groups should have less number of semantic features. For this reason, their competence (all features) can be easily elicited by any experimental task. In this sense, although we have anticipated more MODs in YAK1, the subjects probably exhibited competence rather than the performance reflection.
of MOD. We have expected YAK2 to show no difference in terms of the reliance on MOD and DIS, since the stress values yielded YAK2 proved that YAK2 processed the NL items categorically. This prediction was born out by the data. We have made two more predictions; YAU2 would yielded the greater number of MODs and KOU2, the greater number of DISs. These two predictions were born out by the data. YAU2 showed the significantly greater number of MODs ( $\chi^2=9.8$ ). KOU2 did not reach the statistical significance level ( $\chi^2=0.9259$ ), but supported the prediction in the sense that the subjects adopted the more number of DISs than MODs.

Result 3: Charts 1-4 in 7.0, Appendices present the results obtained among the four groups concerning kowasu. Charts 5-8 in 7.0, Appendices present those with yaburu. Each of the four charts deals with the items belonging to the same MDS cluster. Apart from one exception of item 5 (heart) in yaburu, all the items exhibit zig-zag patterns predicted by the U-shape behavioural growth within the limitation of present coarse unit. Regarding item 5, subjects might have come across somewhere the TL expression 'breaking one's heart', which might have biased the occurrence of the predictable pattern. Or, it could be something to do with the idiosyncracy of this expression.

The present uniform results give some indirect support to Piaget's INCR operations which we used to explain the appearance-disappearance-appearance pattern, by assuming that this pattern is the essential aspect of U-shaped behavioural growth. It is noteworthy that this pattern is at least predicted by the operation of N and C*R.

Result 4: Both concretization and 'abstractization' were observed in the data. The following table present the 'sure' items which are classified into three categories of 'concrete', 'middle' and 'abstract'.
Item 6 (bank-note) in *yaburu* was consistently conceived by all the subjects as the 'sure' item. As we have argued (chap. 4, § 10.1.3, & § 10.1.4), this indicates 'concretization', reflecting the unsuitable syllabus design (see chap. 4 and § 10.1.3). We have hypothesized that this concretization would be counteracted by the process we called 'abstractization'. This prediction was also born out by the data, as Table 18 clearly indicates. This prediction was made on the basis of Piaget's INCR. In this sense, the lexical judgement of translatability appears to give another indirect support for the INCR operations. We have argued in § 10.1.4 that INRC explains accommodation and assimilation. The accommodation process of the TL system involves N or C R, lending to the negation of part of NL system, while the NL assimilation involves I and C. According to the results 3 & 4, the latter case does not show up clearly in the performance of translatability judgements. This suggests that assimilation occurs at the different level. It may be reflected only as the "reversible and reciprocal" relationship between sense and sound or sign (see §3.4). This point needs to be investigated in future.
Chapter 11 Conclusion

1. The present investigation started with my follow-up study of Kellerman's research. Kellerman used an NL transfer paradigm after he obtained the objective representation of NL inter-lexical structure by the learner-oriented method of MDS (see chap.4). This methodology made his NL transfer experiments less speculative than usual. The present research confirms this aspect of his approach. The present research suggests that what we call NL transfer could offer a useful way of examining the process of language learning and acquisition.

In order to make the NL transfer experiment appropriate for the purpose of examining the acquisitional process, we must first of all obtain an NL representation which reflects the learner's internalized linguistic knowledge. For this reason, this methodology requires an accurate measurement of NL meaning. Kellerman adopted MDS, but it appeared that linguistic materials involving discrete or categorical semantic features are not suitable for the dimensional scaling of MDS (see chap.9). This suggested that the MDS distances were not valid for the prediction of translatability judgement. The Thurstone-Torgerson method was more useful in assessing the judgemental limits.

2. Kellerman's research method was, however, useful in that the NL transfer reflects the analogical process. I have tried to develop this aspect of his research. This aim led me to include the operational definition of word-meaning in this thesis (see chap.3). The definition is derived from the psychological experiments concerning concept-formation with the analogical process in mind. I have defined Rosch's PH using distributive lattices, so that I can connect the analogical process with the several kinds of processing. This definition enabled me to relate Piaget's assimilation and accommodation with the judgemental process. Part of the discussion in chap.3, 4 and 10 relates to the issue that the IL hypothesis of restructuring vs recreation can be studied in view of Piaget's assimilation vs accommodation.
3. Corder proposed the three stages of language acquisition consisting of (a) data-processing, (b) hypothesis formation and (c) hypothesis testing. The present research relates to these stages.

(a) As for the data-processing, the six kinds of data processing were distinguished. Of the six, the categorical processing accounted for the data obtained (see chap. 9). This categorical processing was defined, based upon Shepard (1974). The discussion showed that the present definition of word-meaning fit in with the definition of categorical processing. It was argued that the categorical processing involved the parallel operations at least dominant, separable, asymmetric separable and integral processings. In order for a word to be processed properly, it appeared that the four kinds of processing needed to be operative in parallel (§9.3.2 (2)). This assumption of parallel operation was supported by the translatability judgement data. Since there is a clear relationship between the three similarity principles and processings (see Diagram 12, in chap. 10), it was possible to see whether the four kinds of processing operated in parallel. The data appeared to indicate that by employing the three principles of similarity at once, the subjects reorganized the NL items before they arrived at a conclusion. This recurrent NL reorganization based upon the 'feel' of similarity was also shown by free recall experiments in chap. 6. These findings might prove that the categorical processing involves the parallel operations of four processings mentioned above.

(b) As for the formation of hypothesis, the preliminary framework presented in chap. 3 was useful. We have seen that the formation of hypothesis involves two concomitant processes: the abstraction of prototype and acceptance limit (i.e., the limit of application of words). The latter was called here 'judgemental limit'. The Thurstone-Torgerson method was adopted to estimate the judgemental limit, since the experiments in chap. 7 confirmed this application. The analysis in §10.1.4 indicated that the subjects induced the judgemental (inductive) limit
in trial 1 and that the same judgemental limit was used projectively in later trials (and for this reason the latter is called 'projective limit' (see §7.4 and appendix 5).

(c) Chapter 10 is concerned with "hypothesis testing" revealed in the translatability judgement. The youngest group YAC3 and KOC3 showed that the assumption of inductive and projective limit was tenable. The effect of processing upon the judgemental process was also observed. Although all groups essentially adopted categorical processing, performance factors appear to cause the subjects to adopt a specific form of processing to cope with the task (§9.3.2). As a reflection of performance KOC3 adopted separable processing and YAC3, dominant processing. The former relates to distributive lattices and the latter, modular lattices. In accordance with this difference, KOC3 yielded one 'sure' item and YAC3, five 'sure' items, except for the experimentally supplied 'sure' item (see chap.4 for the definition of 'sure' item). Piaget's "vicariance" explained the source of these sure items and they were characterized as 'relative complements'. These analyses indicated that the translatability judgement was a cognitively principled activity.

The translatability judgement among the older subjects also indicated the usefulness of Piaget's approach, i.e., INRC operation. It explained the phenomenon of 'concretization' and the zig-zag patterns of translatability acceptance scores. In this sense, Piaget's framework is worth investigating in relation to IL studies in future.

With regard to the effect of processing upon the judgemental process, we only obtained the general tendency that yaburu showed the heavier reliance upon modular lattices and kowasu, distributive lattices. The categorical processing suggested that the interplay among the four kinds of processing (which constitute the categorical processing) was relative. For this reason, it was not possible to predict a special influence of one or the other type of processing. The experimental demonstration of parallel processing is another future topic of research.

4. When the inference routes were reconstructed, using
the MDS configurations, these routes resembled associative reasoning which has been observed among L₁ children (see chap.3 and 10). In chap.3 we have seen that these associative chains might be synchronic reflections of diachronic changes of meaning. Besides, the judgemental process appeared to have governed by some principles such as similarity principles and lattices, irrespective of the differences of age. This may suggest that the analogical process we have investigated may be regarded as one of the cognitive principles we all possess.

5. The present research tended to give more emphasis to the methodology of NL transfer as a means of examining the L₂ acquisitional process. For this reason, I have only two proposals to make for the classroom situation. (a) It is to teach words "at the basic level" by providing the students with the "concrete images" (see chap.4, 6 & 10). The experiments involving "dominance effect" in chap.6 and 10 along with the undesirable concretization discussed in chap.4 and 10 suggests this proposal. (b) the present experiments showed that the translatability judgements were the outcome of principled cognitive activities. This suggests that the teacher can make use of the learner's translatability judgements on teaching the TL lexes, predicting the 'correct' ones and correcting the predictable erroneous judgements. The teacher's guidances to the students should correspond to the cognitive principles.
Chapter 4

1 On the basis of Garner (1976) and Shepp (1978) the three kinds of processing are illustrated in the main text. The proper definitions by Garner (1974a&b, 1976a&b) are given in §9.2.

2 Sometimes, superordinate terms of natural objects occur with kowasu.

3 & 4 Similar English sentences to (3) and (6) are explained in appendix 2. The TL sentences are "The waves broke on the rocks" and "The tree broke his fall".

5, 6, 7, & 8 (see appendix 2)

9 The sentence illustrates metonymy, since it is the bone of the leg that is broken. This is also briefly discussed in appendix 2.

Chapter 5

1 & 2 It is customary for psychologists to make a distinction between association and imagery. Imagery is defined by Paivio et. al as "a word's capacity to arouse nonverbal images" (1968). Association values are often determined by the number of verbal associations during a specified period of time (e.g., Umemoto, 1969).

3 According to some introspective oral reports given by the subjects, the experimenter's question ("How familiar or how rare your friend's sentences are?") was not understood clearly by the subjects. They appeared to have interpreted the question as meaning how often the events denoted by the sentences could occur in reality. Some subjects report that, unless the events occurred, we do not use these sentences. According to these reports, the subjective familiarity rating itself was not particularly a revealing procedure.

4 The 11 words chosen are not typical in a usual sense. But the typicality condition means that in terms of typicality (in the sense of frequency count) words seen in this characteristic may be compatible to each other. So, although they are rare items, they are equivalent across the age-groups and across the items on the list, only in the sense of frequency count.
As reported in chapter 6, the kowasu sentences are examined in terms of association values. Following closely a standard psychological experiment, the number of associations per sentence (produced by the subjects) was counted. A chi-square test showed no difference.

Although a tree is characterized as undirected, the method makes the subject connect pairs until all subtrees are connected into one tree. Since this tree does not involve crossings (i.e., cycles), it amounts to forcing the subject to draw a line along which N stimuli are arranged. It appears that, although it is defined as undirected, the tree method is a form of directed graph. That is, there is a distinct direction from the most similar (or dissimilar) pair. This suggests that the reference item is implicitly assumed. Thus, the result of the small-scale experiment reported in the main text has some relevance in questioning the additive method.

Luce (1977) examined the assumptions of Thurstone's model in detail in the light of the current advance of studies of the central neural system. In the neural schema there are four modules: a system of parallel nerve fibres, the pulse train, the monitor system and the aggregation system. A particular auditory or visual stimulus activates a system of parallel nerve fibres via transduction, in which the pulse train on a single fibre encodes both intensity and frequency and the information is transmitted in the form of pulse rate. The pulse rate can be regarded as the count of pulses obtained in a fixed period of time or as the time required to complete a fixed count. Then the individual estimates from a substantially identical set of fibres which are monitored by an attention mechanism are aggregated in the central nerve system.

Although specific equations have been suggested for each of the four modules, the system of aggregation in relation to the attentional monitor is now known. Therefore, Luce argues that (1) if the central nerve system takes an averaging procedure, or if the estimates are of comparable magnitude, the subsequent aggregation is a normal distribution (1977: 468); (2) although the system of parallel fibres follows a double exponential function, it does not differ from normal distribution. Third, his own more specific experiments based upon the neural schema above indicate that the aggregated sensory estimates are of the Thurstonian representation. On a final note of caution he says that, magnitude estimation might be more complex, depending on whether the successive stimuli activate the same or different group of fibres.

Since Torgerson's equations follow Thurstone's closely, it is customary to refer to Torgerson's method as the Thurstone-Torgerson method. I have called their method 'SIM' in chapter 4.
Chapter 8

1. The experiment consists of the consecutive learning of two lists and after some retention interval the memory of the first list is tested. RI refers to the detrimental effect of recently acquired information on previously learned material. If the items in the two lists are similar, interference from the newer list results.

2. Lambert summarized the result later in a slightly different light (1969). He classified the subjects according to the acquisitional context into coordinates and compounds and found that there was a significant difference between concordant and discordant lists (p < 0.06). Since in free recall, various associative links are operative, including affective meaning which are found to be different (cf. §1.2), the result does not necessarily support the coordinate and compound dichotomy.

Chapter 9

1. See §3.4 and §10.1.4.
References


Andersen, E. S. "Cups and glasses: Learning that boundaries are vague". Journal of Child Language 1975, 2.


Anderson, R. W. "The impoverished state of cross-sectional morpheme acquisition/accuracy methodology (or: The leftovers are more nourishing than the main course). In Henning, C. (ed.) (1977).


Atkinson, R. C. & Juola, J. F.


"Distinguishing between prototypes: The early acquisition of the meaning of object names" in Kuczaj, S., II (ed.) (1982).


Berlin, B., et. al. (1968) "Covert Categories and folk taxonomies", Ame. Anthropology. 70.


Bialystok, E. "Les premiers stades du language de l'enfant". Journal de Psychologie 1921, 18.

Block,B. "English verb inflection", Language. 24, 7, 1948.

Bloom,L.M. (1973) One word at a time: the use of single word utterances before syntax". Mouton.


405


Bowerman, M.


"The acquisition of word meaning: An investigation into some current conflicts" In Waterson, N. & Show, C. (Eds.) (1978).

"Reorganizational processes in lexical and syntactic development" in Wanner, E. & Gleitman, R. eds. (1982).


Brown, R. W. "How shall a thing be called?". Psychological Review 1968, 65.


"Reference in memorial tribute to Eric Lenneberg". Cognition 1976, 4.


"The acquisition of the English auxiliary by native


Carroll, J. B. (1959) "Review of the measurement of meaning by Osgood et. al", Language, 35.


Chomsky, N.


"Strategies and the Mapping Problem in First Language


Corder, S. Pit.


"Language-learner language" (1978b). In Richard (1978; 71-93)

"Survey article: error analysis, interlanguage and second language acquisition" (1978c). V.Kinsella (ed.) *Language Teaching & Linguistics: Surveys* Centre for Information Language Teaching & Research, C.U.P.

"Language distance and the magnitude of the language learning task" (1979). In Corder (1981).

"Formal simplicity and functional simplification" (1980).
In Corder (1981).


Derwing, B. L. (1973) Transformational grammar as a theory of language acquisition. C.U.P.


Dommergue, J.-Y. & Lane, H. "On Two Independent Sources


Dulay, H. & Burt, M.


Evers, K. (1970). The effects of bilingualism on the recall
of words presented aurally. University of Minnesota, Dissertation Abstracts, 5197 A.


Freedman, D.E.L. "The Acquisition of Grammatical Morphemes by Adult ESL Students". TESOL Quarterly Vol.9, No.4 Dec. 1975. (see Larsen-Freeman, D.)


"The Development of L2 Intuitions". TESOL Quarterly,


Hall, J. F. (1954) "Learning as a function of word
frequency", Amer. J. Psychol. 67.


The state of the art (1968). The Hague.


Hyman, R. & Frost, N.H. "Gradients and schema in pattern recognition". In P. Rabbitt (Ed.)


Indow, T. "Two kinds of multidimensional scaling methods as tools for investigating color space from the macroscopic point of view". Acta Chromatica. 1963, 1.


Indow, T., & Shiose, T. "An application of the method of multidimensional scaling to perception of similarity or difference in color". Japanese Psychological Research, 1956, 3.

Indow, T., & Shiose, T. "A note on an application of the method of multidimensional scaling to perception of similarity or difference in color". Japanese Psychological Research, 1958, 5.

Indow, T., & Uchizono, T. "Multidimensional mapping of


Kellerman, E.


"Transfer and non-transfer: where we are now". Second Language Acquisition 1979, Vol.2. No.1.

"Now you see it, now you don't" in Gass, S. & Selinker, L. eds. (1983).


"An Eye for an Eye: Crosslinguistic Constraints on the


Kemler, D.G. & Smith, L.B.


Komori, T. (1968) "Kioku ni okeru kigo-ka no yakuwari (The functions of coding in memory)." Faculty of Education, Kyoto University, 14.


Kosslyn, S. M. (1978) "Imagery and internal representation". In Rosch and Lloyd (eds.).


Krashen, S.

"Some issues relating to the monitor model" (1977) in On TESOL 1977. Washington: TESOL.


"Relating theory and practice in adult second language acquisition" In Felix, S.W. ed. (1980).


Kruskal, J.B. (1964) "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis". Psychometrika, 29.


Lambert, W. (1969) "Psychological studies of the interdependencies of the bilingual's two languages". In Puhvel (ed.).


Luce, R.D.
"A threshold theory for simple detection experiments". Psychological Review, 1963, 70.


Mandler, G. (1967) "Organization and memory". In Spence & Spence (eds).


Marslen-Wilson, W. D. (1976) "Linguistic descriptions and psychological assumptions in the study of sentence processing". In Wales and Walker (eds).


Marslen-Wilson, W. D. & Welsh, A. (1978) "Processing interactions and lexical access during word recognition in continuous speech", Cognit. Psychol. 10.


McCawley, J. D. "Acquisition models as models of acquisition" in Fasold and Shuy (eds). (1977).


Menyuk, P. & Menn, L. (1979) "Early strategies for the perception and production of words and sounds" in Fletcher & Garman (eds.)


Morton, J.


"Concept, word and sentence: interrelations in acquisition and development". Psychological Review, 1974, 81.


"Experimental Gambits in the Service of Language-Acquisition Theory: From the Fiffin Project to Operation Input Swap" in Kuczaj, S., II (ed.) (1982).


Nelson, K., Rescorla, L., Gruendel, J., & Benedict, H. "Early


Olson, G. M. & Laxar, K. "Asymmetries in processing the terms 'right' and 'left'". Journal of Experimental Psychology, 1973, 100.


Osgood, C. E.


Paivio, A. U., Yuille, J. C., & Madigan, S. A. "Concreteness, imagery and meaningfulness values for 925 nouns". Journal of Experimental Psychology, 1968, 76 (1, Part 2). (b)


427


Logic and psychology (1953). Manchester University Press.


"Facilitation & inhibition in the processing of signals". In Rabbitt & Dornic (1975).

Chronometric explorations of mind: the third Paul M. Fitts lectures (1978). Lawrence Erlbaum Associates,


Posner, M. I. (1975) "Facilitation & Inhibition in the Processing of Signals". In Rabbitt & Dornic (eds).

Postman, L. "Effects of word frequency on acquisition and retention under conditions of free-recall learning".


Quine, W.V.


"Mind and verbal dispositions" in S. Guttenplan et al. (1975).


Reed, S.F., Ernst, G.W., & Banerji, R. "The role of analogy in transfer between similar problem states". Cognitive Psychology, 1974, 6.

Reed, S.K.


"Schemes and theories of pattern recognition" in


Richards, J. C.


Rosch, E.


Selinker, L.


Shepard, R.N.


"The analysis of proximities: multi-dimensional scaling with an unknown distance function Part II" (1962). Psychometrika. 27.


Shepp, B.E. & Swartz, K.B. "Selective attention and the processing of integral and nonintegral dimension: a developmental study". Journal of Experimental Child

433


Swain, M., Dumas, G. & Naiman, N. "Alternatives to Spontaneous Speech: Elicited Translation and Imitation as indicators of second language competeness". Working papers on bilingualism No.3.


Takeuti, G.


technique). Tokyo University Press.


Thurstone,L.L.


"A mental unit of measurement" (1927). Psychol. Rev., 34.

"Three psychological laws" (1927). Psychol. Rev. 34.


The measurement of values (1959). Chicago Univ. Press.


van Buren, P. (1975) "Semantics and language teaching". In Allen & Corder (eds).


"Explorations in semantic theory". In Sebeok (1966).

Werner, H. (1948) *Comparative Psychology of Mental Development*. Follett (Science Edition Inc.)


Winston, P. H. (1977) "Learning to identify toy blockstructures" In Johnson-Laird & Wason (eds.)


Young, G. & Householder, A. S. (1938) "Discussion of a set of points in terms of their mutual distances", Psychometrika, 3.


