Notebook 3

THE PRESENT STATE

OF

THE LOGLAN™ LANGUAGE

by

James Cooke Brown

PREFACE

The present work is a revision and extension of The Institute's two previous notebooks, both published in 1982, and the two special issues of **The Loglanist, TL6/1** (1983) and **TL7/1** (1984), and it incorporates some material from the latter two works. The scope and organization of this present work is quite new, however; in particular it is the first complete description of the language to be published since 1975.

This account of New Loglan is long overdue and I apologize. Both financial reasons and reasons of personal health have slowed The Institute's work down since the early 1980's, when so much seemed to be being accomplished. One of the reasons is that that certain parlor game which had supported the Loglan Project for so much of its life was withdrawn from the market in 1983, and Loglan has had to go it alone ever since. Fortunately, we are now on the verge of Going Public Again; so the long, dry period of Loglan's being everybody's "poor relation" may soon be over. If, as everybody seems to think, Loglan is about to become at least a modest commercial success, the project may at last become financially self-supporting.

I wish to take this occasion to acknowledge the stalwart few who have contributed to the work of The Institute over these last few, difficult years. Faith Rich has made a large contribution to the next Loglan dictionary by completing the Eaton Interface. She was ably helped in doing so by Jeffrey Taylor, Kieran Carroll and Robert McIvor. Unfortunately, their work is not quite ready to be published. But it will, as I say, form the bulk of the next dictionary of the Loglan language whenever that is published.

My daughter Jennifer Fuller Brown managed to bring the Optional Case Tags Project to a happy conclusion this Spring; and the fruits of her work are in this notebook. Paloma Ibañez ably assisted me in bringing the Scientific Borrowings Project very nearly to an algorithmic conclusion; but the fruits of that project are, as explained elsewhere, not quite ready to be published. Glen Haydon has helped me put together the two MacTeach programs that are now available. Bill Greenhood has counselled me from time to time on the proprieties of scientific word-making. And Scott Layson has made yet another extraordinary gift to the project by updating all the Lyces software--which is the tool with which I do my grammatical work--for the more capacious environment of our new Zenith 100 computer.

Users of this notebook are invited to send in (1) notices of whatever errata they may find, and (2) proposals for improving the language by adding to, changing, or deleting any of the provisions described in this notebook. Please keep these two kinds of contributions separate, however. Formal proposals will go to the Loglan Academy for assessment when they meet in the early Spring of 1988; and the format for making proposals formally to the Academy has been described in a recent **Lognet**. Notices of errata should also be kept separate from the covering letter. Preferably they should be on sheets or cards that may be filed separately from correspondence.

We at The Institute look forward to a vigorous testing of the language described in these pages, and to GPA-ing with it in the very near future.

JCB
Gainesville
July 1987

TABLE OF CONTENTS

Foreword Page 11

CHAPTER 1. PHONOLOGY (SOUNDS & SOUND-VARIANTS) 15

1.1 Definitions and Conventions 15
1.2 Two Types of Phonemes 17
1.3 Regular Phonemes 17
1.4 The 6 Regular Vowels 17
1.5 The Advantages of Romance (aa) 18
1.6 The Two Spellings of 'e' of 'met' 19
1.7 The Odd Spellings of /i/ and /y/ 19
1.8 /y/ as a Hyphen 20
1.9 /y/ as a Buffer 20
1.10 /iy/ as a Hyphen in Buffered Dialects 21
1.11 The Effect of Hyphenating and Buffering on Stress 21
1.12 /y/ in Names 21

Table 1.1 Permissible Pronunciations of the Twenty-Five Loglan Vowel-Pairs 22

1.13 The 25 Vowel-Pairs 22
1.14 The 10 Optionally Disyllabic Vowel-Pairs 23
1.15 The Pair-from-the-Left Rule 23
1.16 Indications of Syllabicity 24
1.17 The 17 Regular Consonants 25
1.18 The Odd Sounds of [C c] and [J j] 27
1.19 The Four Vocalic Consonants 27
1.20 The Unfamiliar Consonant Pairs 28
1.21 The Three Irregular Phonemes /q w x/ 28
1.22 The Use of Irregular Phonemes in Names 29
1.23 Three Stress Phonemes 29
1.24 One Pause Phoneme 30
1.25 Intonation 30
1.26 Buffered Dialects 31

CHAPTER 2. MORPHOLOGY (WORDS & WORD-FORMS) 33

2.1 Design Objectives 33
2.2 Definitions and Conventions 33

Table 2.1 The Two Partitions and Three Classes of Loglan Words 35

2.3 Two Major Partitions and Three Word-Classes 35

A. THE MORPHOLOGY OF NAMES 37

2.4 The Forms of Names 37
2.5 The Pause Before Vowel-Initial Names 37
2.6 The Name-Marker Restriction on Names 37
2.7 Working Around the Name-Marker Restriction 38
2.8 Derivations of Names 39
2.9 Internal Names 39
2.10 External Names 39
2.11 Auditorily-Modeled External Names 39
2.12 Visually-Modeled External Names 40
2.13 The Linnaean Polynomials 41
2.14 Pronunciation of the Linnaean Polynomials 42
2.15 Writing Linnaean Names 43
2.16 The Post-Nominal Pause 43
2.17 Resolving Names 44

B. THE MORPHOLOGY OF STRUCTURE WORDS 45

2.18 The Functions of Structure Words 45
2.19 The Four Little-Word Forms 45
2.20 Compound Little Words 46
2.21 Letter-Words 47
2.22 Suffixes for the 52 Latin Letter-Words 47
2.23 Suffixes for the 48 Greek Letter-Words 48
2.24 Uses of Letter-Words 48
2.25 Spelling Aloud 49
2.26 Little Word Predicates 49
2.27 Mathematical Predicates 49
2.28 The "No Pausing Inside Words" Rule 50
2.29 Acronymic Predicates 50
2.30 Pause and Stress Around Acronymic Words and Letter-Words 51
2.31 Pause and Stress Around Dimensioned Numbers 52
2.32 Acronym Recovery Rules 53
2.33 Resolving Structure Words 54

C. THE MORPHOLOGY OF PREDICATES 57

2.34 The Functions of Predicates 57
2.35 A Temporary Stress-Marking Convention 57
2.36 The Post-Emphatic and Intervocalic Pauses 58
2.37 Stress in Predicates 58
2.38 The Forms of Predicates 59
2.39 Three Kinds of Predicates 60
2.40 Primitives 60
2.41 Complexes 60
2.42 Borrowings 61
2.43 Consonant-Pairs 62
2.44 Permissible Medials 62
2.45 Intelligibility at the C/CC-Joint 62
2.46 Hyphenation 63
2.47 Permissible Initials 63
2.48 The Decipherability of Complexes 64
2.49 Affix-Length and Frequency of Use 64
2.50 Term-Reduction 64
2.51 Long Affixes 65
2.52 Short Affixes & Their Derivations 65
2.53 Affix-Assignment & Coverage 66
2.54 Preempted CVr Affixes 66
2.55 Making Complexes 66
2.56 The "Tosmabru Test" 67
2.57 Allowable Borrowings 67
2.58 The "Slinkui Test" 68
2.59 The Resolution & Partial Classification of Predicates 69
2.60 The Predicate Resolution Algorithm 70
2.61 Term-Resolution 70
2.62 The Recognition of Borrowings 71
2.63 Making Borrowed Predicates 71

CHAPTER 3. LEXICON (WORDS & SPEECH PARTS) 73

3.1. Definitions and Conventions 73

(The *-ed lexemes are machine-oriented)

Lexeme A: Afterthought Connectives (Eks) 74
Lexeme ACI: Hyphenating Eks 75
Lexeme AGE: Right-Grouping Eks 75
Lexeme BI: Identity Operators 76

*Lexeme BAD		76
Lexeme CA:	Predicate Word Connectives (Sheks)	76
Lexeme CI:	The Interverbal Hyphen	77
Lexeme CUI:	The Shek Left-Parenthesis	77
Lexeme DA:	Variables	77
Lexeme DIO:	Argument Tags	78
Lexeme DJAN:	Name Words	79
*Lexeme END		79
*Lexeme FI:	The Utterance Ordinal Suffix	79
Lexeme GE:	The Grouping Operator	80
Lexeme GI:	The Fronting Operator	80
Lexeme GO:	The Inversion Operator	80
Lexeme GU:	The Optional Right Boundary Marker ("Comma")	81
Lexeme GUE:	The GE-Group Optional Terminator	81
Lexeme HOI:	The Vocative Marker	81
Lexeme HU:	The Interrogative Argument	82
Lexeme I:	Sentence Connectives (Eesheks)	82
Lexeme ICI:	Hyphenating Eesheks	83
Lexeme IE:	The Identity Interrogative	83
Lexeme IGE:	Right-Grouping Eesheks	84
Lexeme JE:	The First Linking Operator	84
Lexeme JI:	Argument Modification Links	84
Lexeme JIO:	Subordinate Clause Links	84
Lexeme JO:	Metaphorizers	85
Lexeme JUE:	The Second Linking Operator	85
Lexeme KA:	Prefix Members of Forethought Connectives (Keks)	85
Lexeme KI:	Infixes for Forethought Coonectives (Keks)	87
Lexeme KIE:	The Left-Parenthesis	87
Lexeme KIU:	The Right-Parenthesis	87
Lexeme LA:	The Name Operator	88
Lexeme LAE:	Indirect Designation Operators	88
Lexeme LE:	Descriptors	89
Lexeme LEPO:	Event Operators	89
Lexeme LI:	The Left Quotation Operator	90
Lexeme LIE:	The Strong Quotation Operator	90
Lexeme LIO:	The Number Designator	90
Lexeme LIU:	The Single-Word Quotation Operator	91
Lexeme LU:	The Right Quotation Operator	91
*Lexemes M1 through M11		91
Lexeme ME:	The Predifying Operator	92
Lexeme NI:	Quantifiers	92
Lexeme NO:	The Negation Operator	93

*Lexeme NOI: The Negation Suffix 94
Lexeme NU: Conversion Operators 94
Lexeme PA: Inflectors/Adverbs/Prepositions 94

Tense Operators 95
Location Operators 96
Modal Operators 96
Causal Operators 97
The Ga Operator 97

Lexeme PAUSE: The Pause-Comma 98
Lexeme PO: Abstraction Operators 98
Lexeme PREDA: Predicate Words 100
*Lexeme RA: Numerical Predicate Suffixes 101
Lexeme TAI: Letter Variables 101
Lexeme UI: Free Modifiers 103

Attitudinals 104
Discursives 104
Relative Interrogatives 105
Utterance Ordinals 106
Salutations 106
A Note on Other Free Modifiers 106

Lexeme ZE: The Joining Operator 106
*Lexeme ZI: Magnitude Suffixes 107
Lexeme ZO: The Quantity Abstractor 107

CHAPTER 4. GRAMMAR (UTTERANCE FORMS) 109

4.1 Design Objectives 109
4.2 Definitions and Conventions 109
4.3 The Structure of Loglan Grammar 111

Group A. The Optional Punctuators, Rules 1-7 112
Group B. Linked Arguments, Rules 8-19 114
Group C. Predicate Units, Rules 20-33 115
Group D. Descriptive Predicates, Rules 34-48 117
Group E. Sentence Predicates, Rules 49-58 118
Group F. Modifiers, Rules 59-67 120
Group G. Arguments, Rules 68-116 121
Group H. Terms & Term Sets, Rules 117-127 125
Group I. Predicates, Rules 128-154 127

Group J. Sentences, Rules 155-164 130
Group K. Utterances, Rules 165-194 132

LIST 1. PRIMITIVE PREDICATES, L-E (By Loglan) 137
LIST 2. PRIMITIVE PREDICATES, E-L (By English) 147
LIST 3. SHORT AFFIXES 157
LIST 4. CASES OF THE PRIMITIVE ARGUMENTS 167

LIST 5. THE TEACHING CORPUS (LWs are listed where they are first used) 179

A. Imperatives & Responses (eo ao ai ae ti tu mi no) 179

B. Address & Response; Offers & Replies (loi loa sia siu ea oi mu) 179

C. Addressing vs. Naming (la ta e hoi) 180

D. Descriptions (le ne su gu) 181

E. Questions with he; Demonstratives & Plurals; Replacement with da (he da na ri ro levi leva) 182

F. Identity Questions & Sentences; Replacement with de & dui (ie bi hu i de dui) 182

G. Yes/No Questions & Answers; Utterance Demonstratives (ei ia toi toa) 183

H. Tenses; Time Questions & Answers; Local Modification; Punctuation (pa fa ji ipou nahu) 184

I. Time Phrases (pahu fahu fazi pazu tiu) 185

J. Space Questions & Answers; Space Phrases (vi va vu vihu vahu vuhu) 186

K. Existentials & Universals; Completion (ba be bo bu raba rabe rabo nibe ifeu inusoa vina uu) 187

L. Predicate Strings; Grouping, Hyphenation, Connection & Inversion (ge go ci ce ke ki) 188

M. More Connections & Groupings in Predicate Strings (gue cui canoi ka kanoi) 189

N. Event/State Predicates; Other Abstractions (po pu zo di) 191

O. Mass & Event Descriptions; Mass Event Descriptions (lo lovi lepo lopo) 192

P. Specified & Nested Event Descriptions (No new LWs.) 193

Q. Attitude Indication; Conversion, Negation & Superlatives (uo ue ua uu ui nu fu) 194

R. Counting, Quantifying & Numerical Questions (to te fo fe so se vo ve iesu iene ho hoba toba teba foba soba) 196

S. Quantified Descriptions & Questions (iete iefo ieho) 197

T. Measurement, Dimensioned Numbers & Numerical Description (lio lepa -ma -mei -dai) 198

U. Linked Description; Identity Clauses; Replacement with Letter-Words; Mixed Predicates and Arguments (je jue ze sui -mo -ai -ei dai/dei, etc.) 199

V. Identifying vs. Claiming Subordinate Clauses (jio jia) 200

W. Sentence, Predicate & Argument Negation (ni) 201

X. Quotation of Loglan; Fronted Arguments (li lu liu gi) 202

Y. Predicates from Arguments and Prenex Quantifiers (me me- goi) 202

Z. Prenex Quantifiers (goi) 203

AA. Connected Arguments & Predicates; Joint Argument Sets (a anoi onoi noa efa epa gugu do) 204

BB. Causal Inflectors, Modifiers & Phrases (moi soa kou moipa numoi kouhu moihu nukouhu) 206

CC. Compound Term Connectives (enumoi enukou efa eva epa) 207

DD. Connective Questions (ha enoi noenoi) 208

EE. Internal Arguments (No new LWs.) 208

FF. Argument Ordinals (HB-tags) (pua pue pui puo puu) 209

GG. Compound & Connected Tenses (-fa- -pa- -na- ra- ne- ni- -noi-) 211

HH. Logically Connected Clauses (inoca icanoi ica Ice) 211

II. Causally-Connected Clauses (i- ki-) 212

JJ. Indirect Designation; Foreign Quotation (lae sae lie) 212

KK. Metaphor-Marking or "Figurative Quotation" (ja) 213

LL. Letter-Variables and Acronyms (-z-) 214

MM. Predicates as Names & Vocatives (No new LWs.) 215

NN. Grouped & Ungrouped Afterthought Connectives (i- -ge -ci) 215

OO. Spelling (No new LWs) 216

PP. Sentences in VOS Order (goa) 217

FOREWORD

The objectives of the present notebook are three. The first is to provide users of the 1975 language with a description of the present language which will allow them to become competent in Loglan once again. Once that is accomplished, I would hope these rearmed loglanists would then use the enlarged domain of modern Loglan in creative and testing ways, and communicate to The Institute their results. But there is a second objective. I have also tried to create a document that will serve as a teaching text--buttressed, as it now can be, by the two "MacTeach" (computerized flashcard) programs that have recently become available for learning primitives and affixes--but intended mainly for those who wish to learn the current language more or less from scratch. The third objective is to provide a technically complete description of the language that will serve as the easily updated reference manual we will soon need to back up the less formal and more popular publications which The Institute plans to offer to the general public when we go public again...a development of which, we trust, this notebook will be the final forerunner.

Current Loglan has emerged over the last four or five years from the word-making, grammar-expanding, and translating activities of a very few people. Their work has enlarged the language considerably, both in vocabulary, in grammatical domain, and in usage, and is now ready to be reported out. The translating and word-making activities were outgrowths of--actually, they were deliberately undertaken engineering tests of--our more publically-conducted 1976-1982 design studies of usage, grammar, and morphology...the last two having been called affectionately the MacGram ("machine grammar") and the GMR ("Great Morphological Revolution") projects, respectively, while they were still underway.

But well before these various engineering projects had boosted Loglan into a new and higher state, an active corps of competent users, albeit a small one, had developed by 1978 or '79 out of our 1975 publications. To be sure, their competence was in a language--or rather, in what were sometimes highly personal extrapolations from a plan for a language--which was substantially but incompletely described in the two 1975 books, **Loglan 1** and **Loglan 4 & 5**, and in the first four volumes of **The Loglanist**, 1976 through 1980, most notably in the **Supplement to Loglan 1**, a special issue of **TL** published in November 1980 which was the capstone of the first four years of public discussion. Incomplete as those earlier documentations of the language were even then, however, they are now, in addition, very badly out of date. And while there have been two subsequent special issues of **The Loglanist--TL6/1** in 1983 and **TL7/1** in 1984, issues designed to help people catch up with the then-current states of the language--even these two documents largely antedate the recent word-making, translating and grammar-expanding activities and so no longer tell the whole story.

Thus the first goal of the present notebook is simply to update the documentation of the language and make it whole. If that could be done well, I reckoned, then this third notebook would provide a tool with which once-competent loglanists would be able rapidly to restore their competence should they wish to do so. To serve their more sophisticated and often technical purposes, therefore, I have striven mainly to produce a description of the present language as I know it that would be as complete, as technically exact, and as conveniently cross-referenced as I have been able to make it.

It was during the early days of writing for experts in the Spring of 1986 that the notebook acquired its second purpose. A large number of the current partisans of

Loglan, I had been learning, happen to have joined the project well after the creative ferment of the late '70's, and so did not participate in it. Moreover, there are many current loglanists who, although "old hands" in the historical sense, had never actually mastered the old language before it disappeared again into the engineering laboratory. Both kinds of potential users of the notebook began to write me. They, too, hoped to get some mileage out of the new notebook, especially now that developmental research on the language appeared to be slowing down. For these relative newcomers, then, but also for those old hands who have been until now only onlookers, I have tried to erect a second kind of document on the substructure provided by the first. In addition to a technical description of the current language, I have tried to produce a didactically useful, amply-illustrated account of the language from the point of view of the second-language learner. I have tried, in short, to provide these two kinds of sometime students of the language with the means by which, with some personal effort, they may at last become its masters.

These two objectives have not always been easy bedfellows. As the second one began to press itself upon me last Spring I had to admit that a book that promised also to be a reference manual for one-time experts is not an ideal place in which to teach a second language to completely innocent newcomers. Even so, examples are necessary even for experts. And an algorithm or two can be endured by such newcomers as choose to consort with experts. So I have attempted to select the examples and illustrations in this book in such a way that they will, of themselves, constitute a gradual climb through the structure of the language, starting at ground level with the utter simplicity of its phonology, rising through morphological and lexical materials of middle difficulty, and ending with what may, I fear, be found the stratospheric intricacies of the machine grammar. It is I trust a compact account, but it does move through these several levels of intellectual difficulty. (The language itself, of course, remains refreshingly simple...as I trust the reader will soon rediscover. It is just these increasingly exact scientific descriptions of it--which have been made possible and in some sense necessary by our increasingly exact understanding of it--that sometimes border on the intricate.)

There is a third objective of which I have only recently become aware; and that is the possibility that a second edition of **Notebook 3** may even now be looming. Suitably retitled, the next update of this notebook may very well be the one that accompanies the fourth edition of **Loglan 1** to the marketplace. This will probably be in the Spring or Summer of 1988; for it is then that The Institute presently plans to "go public again" with the language. If these plans do indeed develop in this way, then **Notebook 3** may be the first in a long series of continuously updated technical manuals, the purpose of which will be to describe in a single place the current state of the whole language. None of the specific requirements of that looming reference manual have, however, shaped the writing of the notebook...except of course for that ubiquitous canon of completeness, which has been dictated by the first objective as well.

A final note, and an apology. Earlier accounts of the contents planned for **Notebook 3** announced that it would include a small but exemplary vocabulary of scientific borrowings, as well as the algorithm that made them. I meant also to include the translation forays I had made into the international literature of science; for these had provided the test words in the first place and were meant, in the end, to contain them. These translation materials are not included. The latest reasonable date for the publication of this notebook--already twice delayed--was Mid-Summer 1987. I could not make the algorithm for the construction of "best scientific words"--a process that involved, as usual, a statistical analysis of the many judges' opinions I have

collected--in time to include it, and the vocabulary it was intended to make examplary, in the notebook. And to publish my translations with a non-exemplary vocabulary seemed counter-productive. I am sorry to disappoint those readers who expected to find this textual material in this notebook. Perhaps another notebook will be in order after this one has had its day. On the other hand, it seems increasingly likely that the next large task for The Institute, after the loglanists have made whatever use they wish to make of this one, will be GPA (The Institute's acronym for Going Public Again).

But even without the scientific word-list, and the translations that evoked them, the contents of **Notebook 3** will, I trust, be found substantial. It not only contains the most complete description of this developing language that has ever yet been published, it is the first publication since 1975 that even purports to describe the language as a whole. I trust, therefore, that everyone who endures the long march through its lists and pages will have a reasonably good chance of learning to use, for da's own purposes, the extraordinarily rich creative instrument that Loglan has lately become.

JCB
Gainesville
July 1987

CHAPTER 1

PHONOLOGY (SOUNDS & SOUND-VARIANTS)

1.1 Definitions and Conventions: We require a small technical vocabulary. A **phoneme** is a class of one or more speech-sounds all of which are regarded as "instances of the same sound" by the speakers of some language. The individual sounds so classified are called **phones.** All the phones which are members of a given phoneme are called its **allophones.** I shall use strokes [/] to mark off phonemes, sets of phonemes, and phonemic transcriptions. Thus /a/ is a phoneme, /a e i o u/ is a set of phonemes, and /eAmuGODzi/ is a **phonemic transcription.** I shall use parentheses to mark off phones, sets of phones, and phonetic transcriptions. Thus (ah) is a phone, (ah aa) are the two allophones of phoneme /a/, and (ey-AH-moo-GOHD-zee) is the **phonetic transcription** of the utterance phonemically transcribed as /eAmuGODzi/. The utterance itself is **Ea mu godzi;** it means 'Let's go!' Phonetic transcriptions of utterances will sometimes be called **guides;** they are phonetic guides to at least one way of pronouncing those utterances. Phonemic transcriptions are often referred to simply as **productions.** They exhibit one of the ways an utterance might be produced in speech.

Stressed syllables are shown in uppercase letters in both the guides and productions. Hyphens are used to show **syllabification** in guides. Stress rise and fall, shown by the shift from lower to upper case characters, are normally sufficient to show syllabification in productions. But when there are strings of unstressed vowels it is sometimes necessary to show syllable breaks in a production by using **close-commas** (commas without an adjacent space), e.g., /i,aimiGODzi/. This production shows us that the utterance involved was **I ai mi godzi** = 'And yes I intend to go'. The close-comma between them puts /i/ and /ai/ in separate syllables. Without the close-comma, the **default syllabification rule** would pair the vowels from the left. Thus /iaimiGODzi/ syllabifies as (yah-ee-mee-GOHD-zee), and this gives quite a different resolution, namely ***Ia i mi godzi** = 'Certainly; and I go'.

The **asterisk** [*] on ***Ia i mi godzi** shows that this utterance is wrongly pronounced. I shall use leading asterisks [*] to mark all expressions which are illegitimate at some level of correctness, whether that level is phonological, morphological or grammatical. Even bad usages are so marked. ***Ia i mi godzi** is morphologically incorrect because it is missing an obligatory pause before the little word **I**; such pauses are sometimes necessary for resolution. I shall use leading **question-marks** [?] to show that an expression is of unknown legitimacy--unknown, that is, to the reader. Thus **?Ia i mi godzi** is how we would mark this utterance before the reader is expected to know that it is malformed.

Pauses are shown in written specimens either by a **comma-space,** e.g., **Ia, ice mi godzi** ('Certainly; and I go'), or by **period-spaces** which require two following spaces, e.g., **Ia I mi godzi** ('Certainly. I go'). The second marking is equivalent to the full-stop in English. Both pauses are shown in productions by **close-periods,** e.g., /ia.imiGODzi/, and in the guides by **open-periods** (space-period-space). e.g, (yah . ee-mee-GOHD-zee). So the correct pronunciation of these last two specimens is given by the pair of guides (yah . ee-mee-GOHD-zee) for the one that means 'Certainly. I go' and (yah . EE-sheh-mee-GOHD-zee) for the one that means 'Certainly; I go'. A **silence** is the long pause before an utterance begins or after one ends; it is not usually

useful to show silences in productions, but when it is useful to do so they are shown by pound-signs [#]. A **breathgroup** is a pauseless string of phonemes lying between pauses, silences, or pauses and silences in either order.

I shall use **square brackets** to enclose textual **characters**, or sets or strings of characters. Thus [a] and [1] are visual characters; [,] is the comma-space and [.] is the period-space; [.] is the open-period used in guides; while [,] [.] and [#] are the close-comma, close-period and silence, respectively, used in phonemic transcriptions. Although we will rarely have occasion to use bracketed utterances, we can also say that [Ea mu godzi] and [Let us go.] are visually perceived utterances, that is, pieces of text.

Sometimes I shall wish to refer to characters that are not on the fonts available to me for printing this document. I shall say, for example, that in lieu of the proper characters the printed expression [alpha] stands for the Greek lower-case character called alpha while [Alpha] refers to the upper case version of that character.

I shall, as is customary in Institute publications, use **bolding** to mark Loglan specimens whenever their status as text or speech is immaterial. Thus **Ea mu godzi** is the utterance variously represented above. I shall also use bolding to indicate **emphatic stress** in both phonemic transcriptions and guides: /eAmuGODzi/ = /ey-AH-moo-GOHD-zee/ = 'Let's go!' I shall use **single quotes** to mark non-Loglan (usually English) specimens when the mode of delivery is immaterial, and I shall using **underlining** for emphasis in both such non-Loglan specimens and in ordinary text. Thus 'Let's go!' is a translation of **Ea mu godzi** pronounced as /eAmuGODzi/ with no special stress on any word, while 'Let us go!' is a translation of the same utterance when pronounced /eAmuGODzi/ with **godzi** emphasized. (Bolding is also used for titles, technical terms about to be defined, and for chapter, page and section headings in this document...I trust without confusion.) I shall use **double-quotes** for English expressions in the text which are either not being used literally or which have not been defined yet. Thus the allophones of a phoneme may be said to be "instances of the same sound" but are not literally.

The Loglan words which I customarily use as English words when I am addressing loglanists remain unmarked. For example, 'sutori', which means 'at least second' or 'second and subsequent', as in 'the sutori places of the predicate', has become a quasi-English word in my idiolect and so is undistinguished in my English prose...just as the French, German or Latin words and phrases which have been taken into English are now usually unmarked in written English ('That's a priori reasoning' 'He's a gestalt psychologist'). I expect loglandical readers to know the meanings of these borrowings from Loglan, as from German or Latin, or to catch up by looking them up. Besides, they impart a usefully loglandic flavor to one's English thought and speech. The borrowings from Loglan which I notice I use most frequently are 'da' and its kin. I treat these words as genderless, numberless, and caseless 3rd person English pronouns.

I occasionally use the sign '=>' to mean 'is the source of', '<=' to mean 'is derived from', and '=' between a specimen and its translation to mean 'has approximately the same meaning as'. I use parentheses for other uses than Loglan phonetics, of course, as in the next paragraph. I trust this will cause no confusion.

I have used **least equivocal spellings** in ordinary English letters (a procedure made familiar by the Berlitz people) to represent all the Loglan phones we'll need. Thus, I've used (sh) to represent the sound of Loglan /c/, (ee) for the principal sound of Loglan /i/, (oo) for the main value of its /u/, (igh) for the otherwise hopelessly equivocally-spelled 'eye/aye/I'-sound of English, which in Loglan is the sound of the diphthong /ai/, and I've used (eigh) for the sound it invariably spells in such words as

'eight' and 'freight' for one of the two allophones of Loglan /e/; and I have chosen to do this rather than use the special symbols of the International Phonetic Alphabet. The IPA symbols are not widely understood except among linguists; changing fonts to print them would slow down the production of this notebook by about a factor of ten; and our discussion of the sounds of Loglan loses little in technical precision by our adoption of this humbling convention. Phonetic issues in Loglan are not so wide-ranging that they require the full armamentarium of modern phonetic scripts.

1.2 Two Types of Phonemes: There are two types of phonemes in Loglan: the **regular** phonemes that are used throughout the language, and the **irregular** ones that may be used only in (proper) **names.** The term 'name' will always mean a proper name in our discussion of Loglan.

1.3 Regular Phonemes: There are 23 regular phonemes in Loglan: /a b c d e f g h i j k l m n o p r s t u v y z/; so to represent them, Loglan uses the entire English alphabet less the letters [q w x]. Two of these phonemes, /h y/, are new since 1975. /h/ was added in 1981 because it was present in all the target languages and adding it made some much needed word-space available for new little words. Unexpectedly, incorporating /h/ also increased the average recognition scores of Loglan composite primitives (C-Prims) due to the large improvement in the recognizability of the 100 or so primitives that were remade with /h/. Schwa was added in 1981 as the buffer in buffered dialects and was assigned the unused English letter [y]. In 1986 the sound /y/ was given the even more important function of serving as the intraverbal hyphen in regular words; see Sec. 1.8.

1.4 The 6 Regular Vowels: The 6 regular vowels are pronounced as in the following table. The primary allophone is given first, the second, if one is recognized, appears on following lines:

/a/ As in 'father' in the Germanic dialects of Loglan, but as in S. 'casa', F. 'la' and E. 'palm' in the Romance dialects. In the "standard" (Gainesville) dialect, the Romance pronunciation is preferred.

Romance /a/, as in S. 'casa', F. 'la' & E. 'palm'	(aa)
Germanic /a/, as in G. 'vater' & E. 'father'	(ah)

Problems: In unstressed syllables, /a/ is occasionally but incorrectly pronounced as in 'about' or 'sofa', which permits it to be confused with the (uh) sound of /y/; see below.

/e/ As in 'met' (e) or (eh)
Before vowels, the sound of 'eigh' in 'freight' or 'a' in 'ate'; it is also the first part of the diphthong 'ey' in 'grey' (eigh)

Problems: There is a strong tendency for English-speaking persons to use either the monophthong (eigh) as in 'eight' or the diphthong (ay) as in 'day' for /e/ in V-final monosyllables, i.e., to say 'day' for **de** and 'say' for **se**. This must be resisted, if only because Loglan words like **dei** and **sei** also exist in which the diphthong (ay) is genuinely present. This tendency can be resisted by making sure that **de se** are pronounced like the first parts of 'debt' 'set'. These then contrast sharply with **dei sei** as properly pronounced (day say).

/i/	As in 'machine'	(ee)
	Before vowels, 'y' as in 'yet'	(y)
/o/	As in 'note' but shorter; resist the tendency to say 'oh-oo'	(oh)
	Before /r i/, as in 'more' or 'noise'	(aw)
/u/	As in 'lute' but shorter; resist the tendency to say 'oo-oo'	(oo)
	Before vowels, 'w' as in 'woo'	(w)
/y/	'e' as in unstressed 'the', 'a' in 'sofa' and 'above', 'u' in 'up' and 'under'	(uh)

1.5 The Advantages of Romance (aa): Loglan /a/ has two permissible sounds, the Germanic (ah) of 'father' and the Romance (aa) of 'palm' and Spanish 'casa'. I shall usually use the Romance (aa) of 'casa' in the guides simply because I wish to remind the reader that, in the regular words of Loglan (although not necessarily in its names), the Romance pronunciation is preferred over Germanic (ah). Thus, I will usually guide the reader toward the preferred pronunciation (MAAT-maa) of **matma** = 'mother', despite the fact that the Germanic pronunciations (MAHT-mah) and *(MAHT-muh), the second one erroneous, are much more common among current loglanists, who are largely anglophones. (English, along with Dutch, German and the several Scandinavian languages, is a Germanic language.) But remember that you may always substitute the more familar Germanic sound of (ah) for any of my (aa)s. *(MAHT-muh), however, is not an acceptable pronunciation of **matma**. If you're going to use Germanic (ah), you must to do so in both syllables, saying (MAHT-mah) with the (ah) of 'father' in <u>both</u> places.

The chief difficulty with Germanic (ah) is that it is a slack sound that easily degenerates into (uh) in unstressed syllables...especially in the mouths of native speakers of Germanic or Slavic languages. Speakers of such languages regularly and unconsciously allow this slackening to happen to many of their unstressed vowels. So, if you do elect to try to pronounce **matma** consistently as (MAHT-mah), the second (ah) is very likely to slip into a grunted (uh) when you're not listening. Unfortunately, hard as you try, you'll probably end up saying *(MAHT-muh)...which may cause trouble for your human auditors and eventually for your computer. If, on the other hand, you at least <u>try</u> for the tense Spanish (aa) at the outset, trying to say (MAAT-maa) with the stretched lips, wide-open mouth and higher pitch that the production of this Romance

(aa) requires--think of a Spaniard saying 'casa'--then your Loglan /a/ is much less likely to slump off into a grunted (uh) even when it is unstressed.

That's the main reason, then, why you should consider trying to speak the Romance dialect of Loglan: your unstressed /a/s are much more likely to remain intelligible. If you know any Romance languages, you can easily find your models for this (aa). Spanish especially is noteworthy for them. But if you are a monolingual speaker of American English, then the Boston Irish version of 'faather'--if you can conjure it up from a recent movie--is probably your best model...unless, of course, you have The Institute cassettes, which abound in Romance (aa)s.

1.6 The Two Spellings of 'e' of 'met': Notice that the principal allophone of /e/, which is the 'e' of 'met', has two spellings in the pronunciation guides, namely (e) and (eh). These expressions spell exactly the same sound, namely that same 'e' of 'met'. Two spellings are necessary because there is a pair of phonetic contexts in which each spelling would be misleading if it were used in both contexts, yet each is a good guide in one of them. So both spellings are required, each in its own context. Thus (e) is used to guide pronunciation when /e/ just precedes a consonant, as it does in 'met' itself, unless that consonant is /r/. Thus the 'e' of 'met' in **metro** 'meter' is spelled (e) in the guide (MET-roh). But the same 'e' of 'met' when found with following /r/ is spelled (eh), as it is in (KEHR-tee), for example, which is a guide to the pronunciation of **kerti** = 'air'. The same sound is also spelled (eh) when /e/ is final in a word or syllable. This doesn't happen in English, so we can give no English examples. But the 'e' of 'met' is final in many Loglan words. Thus, **ne se pe** all end with the 'e' of 'met'. So the guides for these three words are (neh seh peh), which happen to be the sounds of English 'net set pet' with the 't's omitted.

To take another example, the pronunciation of the Loglan word **ie**--which means 'Which?'--has the value of the first two sounds in English 'ye(t)'. As /e/ is final in **ie** it will be written with (eh) in the pronunciation guide, which is therefore (yeh). Notice that using (e) alone in the guide to **ie**--that is to say, giving its pronunciation as *(ye)--would create quite a different, and for some speakers an erroneous, impression of the sound of this Loglan word. Once again, **ie** is pronounced (yeh), and this is identical to the first two-thirds of the word 'ye(t)'. This is a very un-English way of pronouncing a final /e/, but with a little effort you can learn to do it.

Loglan /e/ has a second allophone, of course. This allophone is the sound of (eigh) in 'eight' which you may have been tempted to give /e/ in words like **ne se pe**, thus pronouncing them like 'nay' 'say' 'pay' incorrectly. But in Loglan, the (eigh) allophone of /e/ is reserved for positions before vowels, and it never occurs anywhere else. Thus (eigh) occurs in words like English 'Mayo' (MEIGH-oh) and Loglan **eo** (EIGH-oh) ('Please') because this /e/ precedes a vowel; but (eigh) is never correct for /e/ when /e/ precedes a consonant or is final in a word or syllable. Native Romance speakers are inclined to misuse Loglan (eigh) in just this way.

1.7 The Odd Spellings of /i/ and /y/: Two of the 6 vowel phonemes will seem oddly-spelled to the English-reading eye, namely /i/ and /y/. These phonemes have been given letters which commonly have quite different phonetic values in English. But the sounds of these Loglan phonemes--which are (ee y) in the case of /i/ and the "grunted" (uh) in the case of /y/--are certainly not odd to the English ear. The sounds (ee y) are in fact very common in English. Indeed, the letter [i] has just these values in

many European languages, including but not only the Romance ones. On the other hand, the short, "grunted" vowel (uh) that is spelled with the letter [y] in Loglan is not consistently represented in any European spelling, the (uh)-sound being usually regarded as a "slighted", or even a "degenerate", version of some other vowel. Thus, the German letter [e] when final is always sounded as (uh), even though [e] has other values in other contexts. In English, (uh) is sometimes spelled with [e] as in 'the' (thuh), sometimes with an [a] as in 'sofa' (SOH-fuh), sometimes with a [u] as in 'upon' (uh-PAHN), and sometimes with an [o] as in 'phonetic' (fuh-NEH-tik); so it does not appear to "have a letter of its own" in English. On the other hand, the sound (uh) is extremely common in English. So having such a familiar sound uniquely represented by the same single letter, even an odd one, may actually be something of a relief for English-reading eyes. Note that the work the letter [y] most commonly does in English is performed in Loglan by another letter, namely [i]. This frees the letter [y] to do its present work in Loglan as (uh).

1.8 /y/ as a Hyphen: As we have seen, the sound of Loglan /y/, which is often called schwa (shvuh) by linguists, is a very short, usually unstressed, "grunted" vowel that occurs in all Germanic and Slavic languages, and in many other consonant-rich languages as well. /y/ has a very odd distribution in Loglan in that in the regular words of **unbuffered dialects**--see the next section for its role in the buffered ones--it occurs only as an **intraverbal hyphen**. In these dialects--which include the standard one--/y/ is the pronunciation of the hyphen-like letter [y] that is used to "glue" the parts of complex words together. An example is **mekykiu** (MEK-uh-kyoo), which means 'ophtalmologist' or 'eye-doctor'. In this word the hyphenated elements /mek/ and /kiu/ are not words, but **affixes**. An affix is a usually shortened representative of one of the words in the **defining metaphor** used to build a complex word; see Sec. 2.55 on Making Complexes. In this case, the defining metaphor is **menki kicmu** (MENG-kee-KEESH-moo), and **mek** and **kiu** are derived from these words by processes that we will consider later; Sec. 2.52. The segment /y/ between them turns them into a single word. The phrase **menki kicmu** also means 'eye doctor' but this time without the English hyphen. When /y/ is used as a hyphen in regular words, it is always a never-stressed sound between affixes. Because the sound (uh) is a hyphen it does not even count as a syllable in locating stress; see Sec. 1.11.

In the regular words of unbuffered dialects the phoneme /y/ has no other use than as a hyphen. That is to say, /y/ is always the sign of the intraverbal hyphen in these dialects whenever it appears in a non-name; see Sec. 1.12 for the uses of schwa in names and below for its uses in buffered dialects.

1.9 /y/ as a Buffer: In the **buffered dialects** of Loglan, /y/ may be used as a buffer between any pair of "difficult" consonants, even when that pair is wholly within an affix or a simple word. For example, in the Japanese dialect of Loglan, the word for 'mother' will probably be buffered. It will become **matyma**, a 3-syllable word pronounced (MAAT-uh-maa). Here the phoneme /y/ is not used as a hyphen. Indeed, **matma** is a simple word without any constituent affixes, so it could hardly contain a hyphen. Instead the /y/ in **matyma** is a buffer between the two "difficult" consonants /t/ and /m/, that is, it makes them pronounceable. (Consonant pairs are **difficult** when the speakers of a buffered dialect choose to buffer them by introducing /y/ between the members of the difficult pair.) If this happens, **matyma** will become a **dialect word**,

that is, a harmless variant of some standard word. It will have exactly the same meaning as the standard word **matma**. Moreover, we expect (MAAT-uh-maa) to be as easily recognized by a speaker of the standard dialect as (MAAT-maa) is likely to be recognized by speakers of the Japanese dialect as a "variant" of their own **matyma**.

1.10 /iy/ as a Hyphen in Buffered Dialects: Notice that the speakers of a buffered dialect may not use /y/ as their intraverbal hyphen. /y/ has already been preempted as their buffer, and to use it for the hyphen as well would lead to confusion for the resolver as to where the "joints" in a complex word actually were. This is contrary to the spirit of Loglan, in which not only are all words uniquely resolvable, but also all terms within words. Using /y/ for both hyphen and buffer in a buffered dialect would make the term-resolver's task impossible in that dialect.

So the hyphen adopted for use in these dialects is /iy/ (pronounced (yuh)). Thus the Japanese speaker of Loglan is obliged to use **mekiykiu** pronounced (MEK-yuh-kyoo) for **mekykiu** (MEK-uh-kyoo) 'ophthalmologist' if da buffers any consonants at all. Of course, if da doesn't buffer da's consonants, da is in that respect speaking the standard dialect. In that case da may use the standard hyphen /y/ without fear of spoiling the resolvability of his speech; see Sec. 2.1 on resolution.

1.11 The Effect of Hyphenating and Buffering on Stress: In the regular words of any dialect the buffer /y/ and the hyphens /y/ and /iy/ are always unstressed. Hyphens and buffers are "so unstressed", in fact, that they do not even count as syllables in locating stress. This rule permits the same syllable to be stressed in both the buffered and the standard versions of a given word, e.g., (MAAT-uh-maa) and (MAAT-maa). Indeed, because of this rule both versions of the word **matma** may be said to be "penultimately stressed"; the penultimate stress rule is discussed in Sec. 2.37. Although **mekykiu** and **mekiykiu** are both stressed on the syllable (MEK) in (MEK-uh-kyoo) and (MEK-yuh-kyoo), they may both be said to be penultimately stressed; for neither hyphen counts as a syllable.

1.12 /y/ in Names: The phonology of /y/ in names is, as usual, a different matter. In names /y/ may occur in either stressed or unstressed positions in all dialects. For example, /y/ appears in a stressed syllable in 'Hunter' => **Hyntr** (HUHN-trr) and in an unstressed one in 'Washington' => **Ua'cintyn** (WAH-sheen-tuhn); see Secs. 2.4-12 on making names.

But what are we going to do with the Japanese version of the name for Amsterdam, which my English-Japanese dictionary lists as 'Amusuterudamu'? Like all loglanists, Japanese loglanists will no doubt be making the effort to pronounce the name of this city in the way its inhabitants do. In standard Loglan, that effort produces **A'msterdam** (AHM-stehr-dahm), for we anglophones who speak standard Loglan are as tolerant of consonant-clustering as the Dutch are. But not so the Japanese. The habit of buffering, which is encouraged elsewhere in their dialect of Loglan, may well cause them to buffer their Loglan names too. In that case, the word for **A'msterdam** in the Japanese dialect of Loglan is likely to come out **A'mysyterydam** (AH-muh-suh-tehr-uh-dahm) or even **A'mysyterydamy** (AH-muh-suh-tehr-uh-dah-muh). Unfortunately this seems to generate two, or even three, different Loglan names for the same place. But the "misfortune" is only apparent. Having one, two or a dozen variant names for

Amsterdam, along with the Loglan standard one with which all the others may be equated by an easy algorithm, is as unlikely to be troublesome for the mutual intelligibility of the two dialects as the fact that buffering creates two words for mothers. Each phonological variant is a synonym of the standard, and once the pattern of buffering in any given dialect has stabilized, each member of such synonym pairs will easily be recaptured from the other. So each dialect word will be as easily understood by the speakers of the standard dialect as the dialect speakers will understand the standard word. Mutual intelligibility of dialects is the aim. We think that consonant-buffering will advance this aim rather than retard it.

Table 1.1 Permissible Pronunciations of the Twenty-Five Loglan Vowel-Pairs

	-a	-e	-i	-o	-u
a-	-- (aa-aa) or (ah-ah)	-- (aa-eh) or (ah-eh)	(igh) --	(ow) --	-- (aa-oo) or (ah-oo)
e-	-- (eigh-aa) or (eigh-ah)	-- (eigh-eh)	(ey) --	-- (eigh-oh)	-- (eigh-oo)
i-	(yaa) (ee-aa) or (ee-ah)	(yeh) (ee-eh)	(yee) (ee-ee)	(yoh) (ee-oh)	(you) (ee-oo)
o-	-- (oh-aa) or (oh-ah)	-- (oh-eh)	(oy) --	-- (oh-oh)	-- (oh-oo)
u-	(waa) (oo-aa) or (oo-ah)	(weh) (oo-eh)	(wee) (oo-ee)	(woe) (oo-oh)	(woo) (oo-oo)

1.13 The 25 Vowel-Pairs: Loglan abounds in vowel-pairs. Not being allowed to enter the syllables of regular words, schwa or /y/ does not figure in any of them. But all possible pairings of the other five vowels /a e i o u/ occur in some words of the language. This creates the 25 vowel-pairs shown in Table 1, which provides a "phonological map" of their pronunciations. The upper entries in each cell of the table are the monosyllabic pronunciations of that particular pair if such a pronunciation is permitted. A '--' is entered on the first line if it is not. The lower entries in each cell are that pair's pronunciation as a disyllable, again if that is permitted.

Notice that some pairs in the table may be pronounced both monosyllabically and disyllabically. Thus /ii/ may be (yee) or (ee-ee). Others, like /aa/, may only be

pronounced disyllabically. Still others, like /ai/, apparently occur only monosyllabically. That is, /ai/ may be pronounced only as the (igh) of **hai** and **sai** and never as *(ah-ee). Incidentally, these two Loglan words are identical in sound to English 'high' and 'sigh'.

In using this table, try to pronounce the pairs with hyphenated guides as two distinct syllables with a "glide" between. It is easier to do this if you stress one of the two syllables relative to the other. While both stress patterns are possible, stressing the first syllable in a vowel-pair is a bit more common in Loglan. As an anglophone, you are likely to find stressing the first syllable a bit more "natural" as well.

We observe that the **always-monosyllabic** pairs are the four "natural" diphthongs /ai ao ei oi/. These can easily be pronounced monosyllabically as (igh ow ay oy) in any context; and so they are.

The next group are the eleven **always-disyllabic** pairs. These are /aa ae au ea ee eo eu oa oe oo ou/. Except for /au/ and /ou/ these pairs cannot be pronounced as diphthongs; and so they never are. /au/ and /ou/ are included among the pairs that are always disyllables to prevent their monosyllabic pronunciations from being confused with one another and with the natural diphthong /ao/ = (ow). So /au/ and /ou/ are always pronounced disyllabically in Loglan, that is as (aa-oo) and (oh-oo) (or (ah-oo) and (oh-oo) if the Germanic /a/ of 'father' is preferred); and /ao/ is always pronounced (ow). This keeps them acoustically quite separate from one another.

The remaining ten pairs are the optional disyllables and are discussed separately in the next section.

1.14 The 10 Optionally Disyllabic Vowel-Pairs: Notice that the 10 vowel-pairs which have both monosyllabic and disyllabic pronunciations in the Table 1 are all those that begin with either /i-/ (**ia ie ii io iu**) or /u-/ (**ua ue ui uo uu**). (In English, we would think of them as the 'y'- and 'w'-words, respectively.) They are therefore quite easy to remember. In certain contexts these optional disyllables are all easily pronounced as monosyllables. For example when they stand alone they are always given the values (yah yeh yee yoh yoo) and (wah weh wee woh woo). But there exist contexts in which these vowel-pairs are difficult to pronounce monosyllabically. Following /r l m n/ seems to be four of these contexts. So in these four contexts, and perhaps some others not yet identified, the loglanist is free to pronounce the i- and u-words either as monosyllables or as disyllables, that is, as (ee-ah) and (oo-ah), etc. Thus, the Loglan word **lui** ('to please...') may be pronounced either as the English name 'Louie' (LOO-ee) or as the French name 'Louis' (l'wee); and which is chosen is largely a matter of personal preference. The word will resolve as **lui** either way. In the Gainesville dialect, the swift French monosyllables seem currently to be preferred.

Uncertainty about the pronunciation of the optional disyllables reflects the fact that The Institute does not know yet how loglanists of a sufficiently wide variety of native-language backgrounds are going to treat them. So we will wait to describe a general pronunciation pattern here, preferring to observe the speech of active loglanists a little longer. In the meantime, no morphological confusion is generated by keeping these options open.

1.15 The Pair-from-the-Left Rule: The default rule for grouping a written string of vowels into pairs is to start pairing them from the left and to restart the pairing process at any marked pause or syllable-break encountered. Once this is done we have to

examine the pairs so made to see whether they are obligatory monosyllables, obligatory disyllables, or optional disyllables. For example, suppose we encounter the written word [aiuiaoea], which is an implausibly long "compound attitudinal"; see Lexeme UI. How do we pronounce it? The first part of the answer is given by the pairing rule. Starting from the left we can try first to syllabify the string as ?/ai,ui,ao,ea/. (We don't know yet whether this is a correct transcription of this new word, so we mark it with leading [?].) We now see that /ui/ is an optional disyllable, and so may be pronounced either (wee) or (oo-ee), and that /ea/ is an obligate disyllable that must be pronounced either (eigh-aa) or (eigh-ah) depending on our personal choice between Romance or Germanic /a/. The other two pairs are the obligate monosyllables (igh) and (ow). We choose to pronounce /ui/ as (wee). This gives us our second transcription, namely ?/ai,ui,ao,e,a/. But what about stress? We decide to use penultimate stress for this new word (which we probably haven't heard yet), for this is another good default rule; see Sec. 2.37. So we produce ?/ai,ui,aoEa/ as our next transcription, this time exploiting the economizing convention that wherever there is a stress-rise or -fall in a transcription, a syllable break is inferrable. Is this the most economical way we can transcribe the pronunciation (igh-wee-ow-EIGH-aa) phonemically? No; it is not. The very pair-from-the-left rule we have been studying allows us to remove the two remaining close-commas. Thus /aiuiaoEa/ transcribes the pronunciation (igh-wee-ow-EIGH-aa) accurately and compactly. It does so because we can infer that there are syllable breaks between the vowel-pairs as counted from the left.

What if someone elects to exercise da's privilege of pronouncing the optional disyllable [u] in the word [aiuiaoea] disyllabically? This rather odd pronunciation--odd because this [ui] is not in the difficult context of leading [l r m n] in which we might expect some readers to prefer the disyllable--can nevertheless be shown by the insertion of two close-commas, one on either side of the second [i]: /aiu,i,aoEa/. This does indeed give (igh-oo-ee-ow-EIGH-aa). One close-comma would not be sufficient here; the pairing rule starts over again with any pause or inferrable syllable break. As a consequence, the transcription /aiu,iaoEa/ would have to be pronounced (igh-oo-yah-oh-EIGH-aa). The interested reader is invited to work out other details of this transcription system.

1.16 Indications of Syllabicity: The novice is of course not expected to remember which pairs of vowel letters in a given string are to be pronounced monosyllabically, or to carry out all the above inferences. So at first we need to tell da. In the prounciation guides the "syllabicity" of a vowel-pair--i.e., whether it is to be pronounced as a monosyllable or as a disyllable--is always plain. All syllable-joints within a breathgroup are indicated by hyphens [-]. So if a pair of vowels is shown with a hyphenated joint, it's a disyllable; if it is not, it's a monosyllable. Thus, (LOO-ee) and (lwee) are plainly the disyllabic and monosyllabic productions, respectively, of **lui**. (They are also the pronunciations of English 'Louie' and French 'Louis'.)

But in phonemic transcriptions, where syllable structure has been played- down so as to lift other features of Loglan speech into prominence, we have seen that certain inferences may be drawn from the transcriptions themselves that will satisfactorily distinguish disyllabic from monosyllabic productions of all vowel-pairs. By way of summary, these inferences proceed as follows.

First, if the two vowels of some pair are shown as having <u>different</u> levels of stress--as they are shown to have in both /eAmuSUCmi/ **Ea mu sucmi** = 'Let's swim' and /muSUCmiEa/ **Mu sucmi ea** = 'We swim, I suggest"--then they are obviously disyllabic in

both productions. Indeed **ea** is one of those eleven vowel-pairs, see Sec. 1.13, that are always disyllabic. So what we are doing in these two transcriptions is showing which of the two vowels in **ea** is being stressed: (eigh-AA-moo-SOOSH-mee) vs. (moo-SOOSH-mee-EIGH-aa).

Second, if both elements of an identifiable vowel-pair are shown as having stress--both emphatic or both normal stress--then that pair of vowels is to be pronounced as a monosyllable. Thus in /AImuSUCmi/ **Ai mu sucmi** = 'Yes, we'll swim' the /AI/ is obviously being pronounced monosyllabically, that is, as the (IGH) of 'high'. Loglan **ai** is the 'Aye' of intention, compliance or consent...as in the nautical 'Aye, aye, sir!'. We can of course write the pronunciation guide for this utterance as (IGH-moo-SOOSH-mee). (But is it possible that you are beginning not to need the guides?)

This leaves pairs of unstressed vowels as the only unexamined case. For example, in /eituMREnu/ = (ay-too-MREH-noo) for **Ei tu mremu** = 'Are you a man?'--literally, 'Eh, you man?'--both phonemes in the interrogative **ei** (which is pronounced, remember, as the English diphthong (ay) in 'say') are shown to be without stress. The interrogative **ei** is often unstressed when it is in utterance-initial position. In utterance-final position, however, **ei** is usually stressed. We can show this by writing /tuMREnuEI/ (too-MRE-noo-AY). We know from Sec. 1.13 that /ei/ is one of the four vowel-pairs that are always pronounced monosyllabically, namely /ai ao ei oi/. But even if we didn't know this, or had forgotten, we could infer the monosyllabicity of **ei** from the transcription /eituMREnu/ itself. On the other hand if we wish to show that two adjacent unstressed vowels are being pronounced disyllabically, as for example if someone spoke the invitation **Ea mu sucmi** without stressing either syllable of the disyllabic /ea/, then we could show that production with a close-comma in the transcription: /e,amuSUCmi/. So we can report the stress possibilities for **ea** in this sentence as the set of three productions: /eAmuSUCmi/ /EamuSUCmi/ and /e,amuSUCmi/. Make sure you can pronounce them all. The fourth logical possibility, /E,AmuSUCmi/, is not really a lively phonological one.

We can now write the third inference as follows:-

Third, if two adjacent unstressed vowels are separated by no mark, and they are indeed members of the same vowel-pair, then they are to be pronounced monosyllabically. Thus, /ei/ in /eituMREnu/ can be inferred to be a monosyllable because no mark separates the unstressed phonemes /e/ and /i/ from each other. (Note that from this point of view, a **stress-rise**, such as occurs in /eA/ (eigh-AA), and a **stress-fall**, such as occurs in /Ea/ (EIGH-aa), are both "marks".) If and only if the two unstressed adjacent vowels are separated by a close comma /,/--as in /e,amuSUCmi/--are they intended to be pronounced disyllabically. Such a close-comma is not a pause; but it does indicate elongation of the time interval occupied by these two unstressed phonemes just as pausing between them would do.

It is important that we transcribe these ambiguous unstressed cases in this explicit way and not rely on the reader's memory of what kind of vowel-pairs they are. Besides, we have those ten optional i- and u-initial pairs to worry about, those for which pronunciation has still not settled down. So we must be especially careful to convey in our guides and transcriptions just which of the several options we are asking the reader to consider.

We have now finished with the Loglan vowel sounds and can address the consonants.

1.17 The 17 Regular Consonants: There are 17 consonants among the **regular phonemes**, which means the sounds used in non-names. Their pronunciations are given

here by English examples. Fortunately, all regular Loglan sounds occur in English, so good examples can always be found. The secondary and tertiary allophones of some phonemes are given on the sutori lines of their entries. The parenthetic expressions in the righthand column of the table are phones. The spellings given in this column are the least equivocal English spellings used in the pronunciation guides:

/b/	as in 'boy'	(b)
/c/	as 'sh' in 'sheep'	(sh)
/d/	as in 'dog'	(d)
/f/	as in 'fat'	(f)
/g/	as in 'goat'	(g)
/h/	as in 'hat'	(h)
/j/	as 'z' in 'azure', '-ge' in 'garage'	(zh)
/k/	as in 'king'	(k)
/l/	as in 'lake'	(l)
	'-le' in 'kettle' when vocalic	(ll)
/m/	as in 'make'	(m)
	as in 'chasm' when vocalic	(mm)
/n/	as in 'net'	(n)
	before /k g/, as in 'sank'	(ng)
	'-ain' in 'certain' when vocalic	(nn)
/p/	as in 'pot'	(p)
/r/	as in 'rot'	(r)
	'-er' in 'father' when vocalic	(rr)
/s/	as in 'sat'	(s)
/t/	as in 'top'	(t)
/v/	as in 'vet'	(v)
/z/	as in 'zoo'	(z)

All phones are single sounds even if the guides to them are written between parentheses with two letters, e.g., (ng). Also, note that the sounds represented by (ll mm nn rr) are what are produced when the normally consonantal continuants (l m n r) are sustained, e.g., as in 'brrr' where (r) is sustained to become the vowel-like (rr). Observe that the sound after (b) in (brr) is long but that it is a single sound. That is to say, no articulatory change occurs within it. When used in this way a consonant is said to be **vocalic.** We shall have more to say on the **vocalic consonants** in Sec. 1.19.

Notice that, except for /c/ and /j/, the letters associated with these 17 Loglan consonants are pronounced exactly as they most commonly are in English. Thus [b d f g h k l m n p r s t v z], while sometimes ambiguous in English, have a most common sound. That most common sound is always at least one of the sounds that letter will have in Loglan. This is true of allophones as well. Thus Loglan /n/ has the three allophones (n ng nn), all of which occur in similar contexts in English. English speakers who repeat the phrase 'sing out for sin gout in the fountain' will hear all three sounds of /n/. The difference between Loglan and English /n/ is that (ng) is one of its allophones in Loglan, whereas /ng/ is phonemic in English. Thus the 'sin/sing' distinction exists in English but not in Loglan. But in both languages the sound that is written as [n] is pronounced (ng) whenever it immediately precedes a /g/ or a /k/. Observe that the English word 'sink' has the sounds of 'sing' in it. The same is true of Loglan. Thus **manko** = 'mouth' is standardly pronounced (MAHNG-koh) in Loglan because the (ng) allophone of /n/ is required before /k/.

1.18 The Odd Sounds of <C c$ and <J j$: Only the letters used to represent /c/ = (sh) and /j/ = (zh) seem to be oddly paired with their sounds in Loglan. The letters [c j] have been assigned to these two phonemes, first, because we need these sounds in this international language--(sh zh) are especially common in Chinese, for example--and, second, because no "digraphs" are permitted in Loglan. (A **digraph** is a sequence of two letters used to represent a single sound; thus [ck kn ng ph pn sh th] are all English digraphs.) But the Loglan rule is that each simple sound must be represented in written Loglan by a single letter. The letter [c] happened to be unemployed. [k] and [s] were doing all its (English) work in Loglan; and [c] is at least weakly associated with the (sh) sound in other languages. For example, French 'chat' is (shaa) and Italian 'ciao' is (chow). It turns out that the (ch)-sound in this Italian word may be further analyzed as (t) + (sh). So (ch) is in fact written in Loglan phonetics as (tsh), and phonemically this is of course /tc/. Thus, with [c] given the value (sh), the Italian word 'ciao' can now be written in Loglan phonemics as /tcao/, and English 'chew' as /tcu/.

Loglan /j/ is parallel to Loglan /c/. Thus (zh) is the **voiced**, that is, the unwhispered, version of **unvoiced** or whispered (sh), and it occurs commonly in French as the French /j/ of 'Jean' (zhaa^{n}). (We shall show the nasalization of a preceding vowel by superscript ^{n}.) (zh) also occurs in certain words of French origin in English, such as 'azure' and 'garage'. But (zh) is also one component of a very common consonant of English, one that is also represented by the letter [j]. For just as the (ch) sound may be analyzed as (t) + (sh) in Loglan, so the sound of English [j] may be analyzed as (d) + (zh). Thus English 'joke' may be written (dzhohk) in Loglan phonetics, and /djok/ in Loglan phonemics. By a precise parallel, 'choke' may be written (tshohk) phonetically and /tcok/ phonemically. This explains why the Loglan word for 'chain' is written **tcena** and that for 'judge' is written **djudi.** (Try to pronounce these as (CHEH-naa) and (JOO-dee), not as (CHEIGH-naa) and (JUH-dee).)

1.19 The Four Vocalic Consonants: The vocalic values of the four consonants /l m n r/, namely (ll mm nn rr), are virtually confined to borrowed names in Loglan. Examples are English 'Earl' which is respelled in Loglan as **Rl** but is still pronounced (RR-ll); 'Burton' => **Brtn** pronounced (BRR-tnn) or sometimes (BRR-tyn); and 'Herbert' => **Hrbrt** pronounced (HRR-brrt). In such English names, as in similar names in other consonant-rich languages, these four **vocalic consonants** play the roles of vowels. So they have vocalic values in the Loglan loan-words **Rl, Brtn** and **Hrbrt** which are derived from these natural names, and which thus require no other vowels.

These same four consonants, however, are seldom pronounced as vowels when they occur in the regular words of Loglan, i.e., in its non-names. Thus the instances of /r/ in the predicates **rodlu** (ROHD-loo) 'road', **farfu** (FAAR-foo) 'father', and **brudi** (BROO-dee) all have consonantal values. In contexts like **mrenu** 'man', however, the /r/ is sometimes vocalic, especially in the speech of newcomers to the langauge. That is, **mrenu** is sometimes pronounced as the trisyllable (mrr-EN-oo) rather than as the disyllable (MREH-noo). But the word is still spelled **mrenu** and transcribed as /MREnu/ in Loglan phonemics; for in regular words (rr) is simply an allophone of /r/. That is to say, (rr) is a legitimate, contextual variation of /r/ but does not have separate phonemic status in the language.

What this means is that hearing (mrr-EN-oo) rather than (MRE-noo) does not cause

a loglanist to think da's heard a different word; but only that da is listening to someone who is not comfortable with this Loglan consonant combination yet.

1.20 The Unfamiliar Consonant Pairs: /mr/ is just one of a handful of Loglan consonant combinations which will be unfamiliar to monolingual anglophones. Some others are /dz/ as in **dzoru** (DZAW-roo) = 'walk'; /ts/ as in **tsero** (TSEH-roh) = 'error'; /cl/ as in **cluva** (SHLOO-vaa) = 'love'; /ct/ as in **ctifu** (SHTEE-foo) = 'stuff/matter/material'; /ck/ as in **ckozu** (SHKOH-zoo) = 'cause'; /cm/ as in **cmeni** (SHME-nee) = 'money'. Despite their curious appearance, none of these consonant-pairs will be difficult for an anglophone to pronounce.

The two pairs /dj/ and /tc/ look difficult but are not. The sounds they represent are actually very common in English, being usually spelled [j] and [ch] as discussed above.

1.21 The Three Irregular Phonemes /q w x/: These three sounds--one of them, /w/, a vowel, and the other two, /q x/, being consonants--occur only in names and letter-words, and then only rarely. All three are, by world standards, "difficult" phonemes. That is to say, each is found in only a small subset of the world's languages and is therefore unfamiliar to most human ears and tongues. But the letters assigned to these three sounds in Loglan, namely [Q q W w X x], occur as **letter-words** (see Sec. 2.21) or as parts of **acronyms** (2.29) in many scientific contexts--for example, the chemical symbol for tungsten is 'W', and the 'X' in 'X-ray' must somehow be accommodated--and they are also employed, although not frequently, in writing one of the most extensive vocabularies of science, namely the Linean binomials of biology. These two-term species names, like '**Homo sapiens**' and '**Escherichia coli**', are by international convention always either italicized or bolded in text and spelled identically in all languages...even in Chinese, Japanese, Hindi, and Russian, languages which do not use the Latin alphabet in any other context. It is therefore advisable that Loglan, too, provide the means, at least, for the transcription of these binomials, and in ways that preserve as much as possible of their standard visual character. To this end, once the 23 regular Loglan sounds had been given letters, the 3 letters left over from the standard Western European alphabet were given to these three fairly commonly used European sounds. Their phonetic values are as follows:

/q/	'th' as in E. 'thin' (unvoiced)	(th)
	'th' as in E. 'then' (voiced)	(dh)
/w/	'eu' as in F. 'bleu', umlaut 'u' in G. 'München'	(eu)
/x/	'ch' as in G. 'Bach', 'k' in R. 'Kruschev'	(kh)

As an anglophone, you will know the sounds (th) and (dh), of course; both are spelled in English with [th]. But you may balk for awhile at spelling them with [Q q]. One useful mnemonic is that upper case [Q] looks quite a lot like upper case Greek theta [Θ] but with its bar slipped down and tipped a little, whence [Q]. Theta is, of course, the letter used in Greek for its (th) sound. The other two irregular sounds are

non-English, but probably familiar enough to those linguaphiles who make an effort to pronounce foreign words correctly.

1.22 The Use of Irregular Phonemes in Names: As suggested by the examples given above, the three irregular sounds of Loglan may also be used, at the name-maker's discretion, to produce better imitations of natural place and person names than would otherwise be possible. Thus German 'Bach' may now be imported into Loglan without change as **Bax** (bahkh). Similarly, 'München' (Munich) may be exactly reproduced as **Mwnxen** (MEUN-khen). The sounds of English 'Theodore' may be exactly reproduced as **Qi'ydor** (THEE-uh-dohr), and 'Kruschev' may be well-approximated as **Xrucyf** (KHROO-shuhf). On the other hand, if the word-maker is concerned about minimizing pronunciation difficulties for the large number of persons, on a world-wide basis, to whom one or more of these three irregular sounds is completely alien, and if the respelling of the natural name in regular Loglan sounds is both sufficiently distinct from other names and a recognizable transformation of its source word, the builder may opt for less than perfect imitations of the natural language sources and use regular sounds entirely. This would give **Bak** (bahk), **Muncen** (MOON-shen), **Tiador** (TYAH-dohr), and **Krucyf** (KROO-shuhf) for the four natural names in question; and at present this is perfectly acceptable.

At the moment, a mixed strategy seems to be most attractive. Thus with the arrival of /x/, 'Bach' is now being enthusiastically rendered by **Bax** and 'Kruschev' by **Xrucyf**. On the other hand, **Tiador** and **Muncen** still seem to be acceptable variants of 'Theodore' and 'München' which avoid the formidable oddities of **Qi'ydor** and **Mwnxen**. It may be that these oddities will turn out to be mainly visual. In that case, they are likely to become less objectionable with the passage of time. So the policy of whether or not to use the irregular sounds of Loglan in contriving the Loglan forms of imported natural names is still an open one. It is likely to remain so for some time.

1.23 Three Stress Phonemes: We have already indicated how stress is written in the guides and productions, namely **no stress** by lower-case letters, **normal syllable stress** by upper-case letters, and **emphatic stress** by both bolding and upper-casing the normally stressed syllable of the emphatically stressed word. It remains to mention that these three values are sufficent to describe all meaningful differences in Loglan stress contours.

For example, the pronunciation guide for the Loglan equivalent of 'You <u>too</u> came!' is

(too-**SWEE** . paa-KAAM-laa)

The Loglan sentence whose production is here being guided is of course **Tu sui pa kamla.** Word-for-word this means 'You also before come'. As we will see in the chapter on grammar, the Loglan tense operator is a separate word; it precedes the expression to be tensed.

Acoustically, stress is an increase in both the amplitude (loudness) and the duration of the sounds that make up a stressed syllable or syllables. Both are measured relative to the other sounds of its utterance. Thus (**SWEE**) is not only the loudest syllable in the above utterance, it is also the longest.

1.24 One Pause Phoneme: The pauses that occur in Loglan, as in any language, vary markedly in both length and significance: from long hesitations to brief phrasing pauses. Nevertheless, for morphological and grammatical purposes, it is sufficient to classify all Loglan pauses as members of a single phoneme. This makes the Loglan pause phoneme a single one with many allophones. For example, the glottal stop in Loglan is an allophone of pause. So those instances of exceedingly brief glottal stops that occur, for instance, between the vowels of a disyllabically pronounced vowel-pair are not counted as pauses at all. In short, as in most languages, Loglan pauses are those silences that occur in the midst of speech that are perceptible as such. In acoustic practice this will mean that there is a number of milliseconds above which a pause, say a glottal stop, is perceptible by most auditors as such and below which it is usually not.

A pause is always marked by a period, or full stop, in the guides and productions. In the guides, open-periods [.] are used; in the productions close-periods [.] are used. The silences that precede and follow utterances are not pauses; they are either not marked at all or marked by the pound-sign [#] in the productions. An utterance so marked is an **isolated utterance.** Thus

/#tu**SUI**.paKAMla#/

is the full phonemic transcription of one way in which the utterance **Tu sui pa kamla** may be produced, namely as an isolated utterance with emphatic **sui** and an optional pause after it. Contrast the phonemic transcription of this production with the phonetic quide to the pronunciation of the same utterance in the previous section. Notice that the phonetic detail has disappeared, leaving the stress/pause "contour" of the production plainly revealed. Because stress and pause together describe a pattern that overlies the whole utterance, the stress and pause phonemes of a language are often called its **suprasegmental** phonemes. Its vowels and consonants are called its **segmental** phonemes.)

1.25 Intonation: Intonation is not a phonemic quality of Loglan speech, and so is not indicated in our phonemic transcriptions. Intonation in Loglan is in fact free to vary in any way a speaker or a dialect group finds natural. It is true that in the Loglan speech we have observed so far questions are generally accompanied by rising tone, and the approaching end of a declaration by falling tone. But this may only be because the Loglan speakers we have observed so far are native speakers of English; and in English this is what happens. Still, in deference to the tonal habits of this potentially "ancestral" group of speakers, a questioning or exclamatory intonation is sometimes shown in the pronunciation guides by the use of [?] or [!].

It will be interesting to see how the tonal dimension of Loglan develops as speakers become more fluent. Will it only reflect the dialectical origins of different groups of speakers in different natural languages? Or will other, language-wide tonal features eventually develop? If so, will these be the result of "universal mechanisms", that is to say, of mechanisms which are inferrably fixed in the human gene pool and so operate in human languages generally? Or will they be only "founder effects", that is, the accidental consequences of the fact that the language started in a certain place--Gainesville, Florida, U.S.A., to be exact--and was first taught by a certain retired college professor who happened to be a native speaker of English, but who also happened

to be foreign-born, to have lived many years abroad, and to have a small acquaintance with certain other European languages? Complete tonelessness, such as often characterizes the speech of badly programmed computers, is probably not even a possibility for human speakers.

Incidentally, toneless computer speech--of the sort that can now be produced by Vocoders and other devices--is surprisingly intelligible when the language is Loglan. This is probably because all the syntactical work normally done by tones in human languages is done by particles in Loglan. Thus, we use the question particle **ei** (ay) (as in 'lay') to transform the declarative sentence **Tu mrenu** /tuMREnu/ ('You are a man') into the question **Tu mrenu ei** /tuMREnuEI/ ('Are you a man?'). (**Ei tu mrenu** /eituMREnu/ works just as well of course, and is slightly to be preferred grammatically.) Notice that the phonetic difference between the two English utterances is far greater than that between the two Loglan ones, and that the English difference is largely in the word-order and intonation. In Loglan, our recognition of the fact that the second utterance is a question could be made to depend entirely on our hearing and understanding the extra word **ei**. So both questions /tuMREnuEI/ and /eituMREnu/ may be spoken quite tonelessly and still be recognized as questions by a loglanist...or by a suitably programmed loglaphone computer.

1.26. Buffered Dialects: Anticipating that many who will eventually speak Loglan will find some of the consonant clusters in its standard, anglophone dialect awkward to produce, we expect the sixth vowel, schwa or what is now Loglan /y/, to be used for "buffering" any or all of those clusters: that is, to separate difficult consonant pairs by interposing this neutral vowel between them. Dialects which used /y/ in this way would then be known as the "buffered" dialects of Loglan.

It is likely that the Japanese dialect of Loglan will be buffered. While Japanese has some consonant pairs, they together comprise only a small fraction of those that occur in standard Loglan. Here's a partial list of permissible consonant-pairs in Japanese: **dj tc ts; n** followed by **c d g k r s t z** and by **dj tc ts; mb mp**. Nearly any other consonant-pair is going to be awkward for a Japanese-speaking loglanist to produce.

As an example of what the Japanese already do about such matters, here is their rendering of Dutch 'Amsterdam': 'Amusuterudamu'. [u] is obviously the letter by which they represent their buffering vowel in Latin transcription. But phonetically the Japanese buffering vowel has a value much closer to schwa /y/ than to our Loglan /u/, being approximately the sound of English 'u' in 'full'. Thus a Japanese dialect of Loglan could well employ /y/, or some suitably neutral local allophone of it, to stand between any pair of consonants that was not on some official list of "pronounceable pairs", a completion of the list begun above. The result would be easy for a Japanese to produce and yet be fully understandable to speakers of the standard, European dialect of Loglan...as 'Amusuterudamu' is understandable to us now. Thus with a little practice in listening to them, **ma'tyma, fa'ryfu, so'ryme,** and **byru'di** would soon be intelligible Loglan words to English-speaking loglanists, whether in speech or in print. Many predicate words would be the same in both dialects, e.g., **ti'tci, da'mpa** and **tse'ro;** and of course none of the Loglan little word sequences--which offer few if any pronunciation difficulties to the Japanese--would require buffering; and so this extensive region of the language would also remain substantially invariant across dialects.

The question arises: If we are going to use /y/ for buffering consonants in the

buffered dialects, what are we going to use as the intraverbal hyphen in those dialects? The two functions must be kept separate; to let one operator serve them both would be to give wrong information about term-joints to the resolver. /iy/ pronounced (yuh) has suggested itself as the hyphenating
syllable to be used in buffered dialects; see Sec. 1.10. /iy/ is easy to pronounce, phonemically similar to /y/, and can easily--one might say harmlessly--be used in all positions in which /y/ is used in the standard dialect. As an example of a word that would then contain both /y/ and /iy/, take **sanpyse'nsi**, the new Loglan word for 'semiotics' in the standard dialect. The consonant pair /ns/ is permissible in Japanese and needs no buffering; but /np/ is not, and so requires buffering. Replacing /y/ with /iy/ and buffering /np/ with /y/ produces **sanypiyse'nsi** (saan-uhp-yuh-SEN-see), an eminently pronounceable word for 'semiotics' in Japanese Loglan...and one which is recognizably the "same" as (saan-puh-SEN-see) in standard Loglan.

CHAPTER 2

MORPHOLOGY (WORDS & WORD-FORMS)

2.1 Design Objectives: The principal objective to be met in designing Loglan morphology was to ensure the **audio-visual isomorphism** of the language. By this is meant that the written and spoken forms of every well-formed Loglan utterance, as those forms would be seen and heard by their readers and listeners, respectively, would in all non-trivial respects be reciprocally inferrable from one another. In short, that the spoken string /la.ARqr.paceNOInaBRAgai/ be convertible always and everywhere into the written string **La Arqr, pacenoina bragai** ('Arthur is no longer king'), and vice versa. It was hoped that this feature would contribute not only to Loglan's utility as a man-machine interface language, where the contribution of isomorphism is obvious, but also to its learnability as a second language by adults, in which the transition from writing to speaking is often a difficult one, and finally also to its acquisition as a second tongue by children, in which the reverse transition is often the difficult one.

Perhaps the most important task to be performed in insuring audio-visual isomorphism is that of **word resolution**, a task of the listener. For the listener to perform this task as swiftly as speech is uttered, and also impeccably, there must be a reasonably simple resolving algorithm--learnable by humans and programmable for computers--that is capable of transforming any heard utterance which has been grammatically composed of well-formed but largely pauselessly-spoken Loglan words, into just one string of properly spaced, capitalized and punctuated written words. Loglan morphology makes such an algorithm both easy to write and to execute in real time. Even a small computer can execute it routinely.

We must note one exception to Loglan's audio-visual isomorphism. If an utterance being read aloud from text involves a Linnaean binomial, that is, one of the standardly spelled pairs of genus and species names which are universally used as naming terminology in biological science (e.g., **Australopithecus afarensis** and **Escherichia coli**), then that utterance will not always be simultaneously transcribable in a form that matches the original text in all particulars. The Linnaean binomials are regularly pronounceable in Loglan phonology; see Sec. 2.13. As such they may be regularly resolved and transcribed; but those transcriptions will not be infallible reproductions of the original Linnaean words. To recapture those originals exactly, thus completing the isomorphism, will require in most cases a tabular lookup that will probably not be executable in real time.

2.2 Definitions and Conventions: First, we need a small technical vocabulary and a number of expository conventions. A **morph** is a simple word or a simple component of a larger word that has a distinct meaning or makes a distinct contribution to the meaning of the embedding word. A morph, in short, is the smallest meaning-bearing component of speech. For example, **pa** is both a simple word meaning (roughly) 'previously' and a component of compounds like **pacenoina**, where it means 'before'. A **morpheme** is a set of one or more morphs with the same or related meanings. If a morpheme has two or more morphs as members, they are said to be its **allomorphs**. For example, both **no** and **-noi(-)** convey the meaning of negation in Loglan ('no' 'not' 'non-' etc.), and the second morph is plainly a derivative of the first; so each is an allomorph

of the morpheme **no**. A morpheme is named for its **free allomorph**, i.e., the one that is derivationally most basic and also capable of functioning as a simple word. In Loglan, these are always the same morph; and in this case it is **no**. Morphs that are not free are **bound**. Thus **-noi(-)** is the **bound allopmorph** of **no** because it occurs only in compounds. The leading '-' in the expression '-noi(-)' means that **-noi(-)** is never initial in its embedding word; the trailing '(-)' means that **-noi(-)** is sometimes medial and sometimes final. In an alternative terminology, the markings on **-noi(-)** show that it is either an **infix** or a **suffix** but never a **prefix**. In this terminology a bound morph is called an **affix**.

A **word** is a string of one or morphs that is treated as a single element by the grammar, just as 'nevertheless' is treated as a single element by English grammar. Words are **simple** if they contain exactly one morph, and **compound** otherwise. Boundaries between words are called **junctures**. It is the business of the **resolver**--a computer program or some subsystem of the human central nervous system--to discover the junctures in a string of pauselessly pronounced words. When it is important to show junctures, we will represent them by the boldface vertical bar [|]. The joints between morphs in a compound or complex word will be shown by the boldface slash [/].

A set of morphs or morphemes whose phonological constituents may be specified in a relatively compact formula is said to be a **form-class** or simply a **form**. We shall use the following conventions to represent the constituents of such forms: 'X' will represent any (segmental) phoneme, that is, any vowel or consonant including the irregular ones /q w x/ and /y/. 'V' will represent any of the **primary vowels** /a e i o u/ or any vowel except /y w/. 'C' will represent any **regular consonant**, that is, any consonant except /q x/. If it is unimportant whether a pair of primaries is pronounced monosyllabically or not, we shall write its formula as 'VV'; if an obligatory monosyllable is called for, we shall write it 'vv'; if we require an obligate disyllable, we shall write 'V,V'.

A **permissible initial** consonant pair is a pair of regular consonants that may be initial in a morph, i.e., in an affix or a word; a **permissible medial** is a pair of consonants that may be adjacent in a morph. Clearly, the permissible initials are a proper subset of the permissible medials. When it is unimportant whether a pair of adjacent consonants is a permissible initial, we shall write it as 'CC'. When a consonant pair in a given formula must be a permissible initial, we will write it 'cc'.

As before, we shall mark proscribed words or forms with a leading asterisk; for example, ***spai** is not a permissible word because *ccvv is not a permissible form. In a demonstration sequence, if the reader is not expected to know yet whether a given form is acceptable, for example, ?ccV,V, it will be marked by a leading question-mark. **?Spea** pronounced (SPEIGH-aa) is one of the instances of this questioned form. Actually, this form too is unacceptable...for reasons that we will find out later.

Formulas and their ingredients are usually distinctive enough to require no special marking. Thus VV is the formula for the 25 possible pairs of the 5 primary vowels. **Silences, pauses, syllable breaks** and **stresses** on preceding vowels or vowel pairs are shown in the formulas by their phonological marks [# . , ']. If a specific phoneme is required in some formula, it will be shown in the formula by the same lower-case letteral that represents it phonemically. Thus the formula dV is the set of 5 morphs generated by putting each of the 5 primaries after phoneme /d/, in particular, it is the set **da de di do du**. The ambiguity this convention seems to create between 'vv', which could mean a pair of /v/s but doesn't, and 'vv', a monosyllabic vowel-pair, is only apparent. If it were a pair of **v**'s, they would be bolded, stroked or bracketed. Besides, "double consonants", or adjacent instances of the same consonant, do not exist

in Loglan. This renders 'cc' unambiguous as well.

To specify a set of equally permissible alternatives in a formula we shall separate them by strokes and enclose the set in square brackets. Thus, [CCV'/CV'C] says that either of these two kinds of stressed syllables is acceptable here. The order in which the alternatives are listed is unimportant. A plus-sign is used as a sign of concatenation between the parts of a word. Thus [CCV'/CV'C]+CV is the complete formula for the form-class to which both **brudi** and **matma** belong. (They are pronounced (BROO-dee) and (MAAT-maa) and mean 'brother' and 'mother'.) Parts of a form which may be iterated one or more times--that is to say, may be present in a larger form as a string of one or more instances of that part--are shown in parentheses in morphological formulas. Thus, (X)+[C/q/x]+. is a formula for the form composed of a string of one or more consonants and vowels of any sort and in any order followed by either a regular consonant or one of the two irregular consonants /q x/, the whole followed by a pause.

Table 2.1 The Two Partitions and Three Classes of Loglan Words

Feature	Irregular Words = Names	Regular Words	
		Structure Words	Predicates
Ending	Always C + /./	Always V	
/q w x/	May	Never	
Stress	Any	Penultimate or none	Always penultimate
/y/	May	Never	As hyphen only
CC	May	No	At least one

2.3 Two Major Partitions and Three Word-Classes: As Table 2.1 shows, the two major partitions of Loglan morphology are (1) between **regular** and **irregular** words, and, among regular words, (2) between **structure words** and **predicates.** There is only one kind of irregular words, namely **names.** Thus the two morphological partitions generate three major word classes: (i) names, (ii) structure words, and (iii) predicates.

Names are the relatively permanent designations of places, persons and things, which, in any given context, are always meant to be unique designations. In Loglan

most but not all names are borrowed from the natural languages. The **structure words** of any language establish the grammatical structure of its utterances and are generally quite short. They tend to be "little words", like 'the' and 'now' in English and **le** and **na** in Loglan, but structure words may also be compounds of short words, like 'nevertheless' in English and **lena** = 'the-present...' in Loglan. **Compound structure words** may sometimes be quite long, for example, **pacenoina** = 'before-and-not-now', or 'no longer'. Occasionally we will speak of simple structure words as **little words.** The words in the third class, **predicates,** are the words of reference in any language. For example, predicates are the nouns, verbs and adjectives of English. Unlike English predicates Loglan predicates are grammatically undifferentiated and have a simple, readily identifiable form.

Table 2.1 gives the features that form the two partitions and identify and characterize the three classes of words. The first two features differentiate irregular from regular words. All irregular words are names, and these and only these may contain the irregular sounds /q w x/. Furthermore, only names end with consonants followed by an obligatory pause; all other Loglan words end in vowels. The next feature, stress, also helps to differentiate names from non-names. Stress may occur anywhere in a name, and if stress in a name is "irregular", that is, if it falls on any syllable except the penultimate one, the stressed vowel is marked with an apostrophe in the written form, e.g., **Pari's.** The next feature also helps to differentiate names from non-names. The grunted vowel schwa or /y/ may occur anywhere in a name and even be stressed, e.g., **Hyntr** (HUHN-trr); but it does not occur in structure words at all and occurs only as a hyphen in predicates, e.g., **mekykiu** (MEK-uh-kyoo). The final feature, the presence or absence of consonant-pairs, differentiates the two kinds of regular words from one another. Structure words never contain adjacent consonants; predicates always contain at least one pair. Names, in keeping with their general flexibility, may or may not contain adjacent consonants.

Because of these deep morphological "furrows" in the vocabulary of Loglan, the learner can tell the broad morphological class of each new Loglan word da encounters at a glance. Take the sentence **La Arqr, pacenoina bragai** (laa . AAR-thrr . paa-ceh-NOY-naa-BRAA-gigh). **La** must be a little word; it is V-final and contains no CC. **Arqr** must be a name; that it is C-final is sufficient to tell us this, but there is another clue. **Arqr** also contains the irregular phoneme /q/ (th). **Pacenoina,** although clearly too long to be a "little" structure word, must nevertheless be a compound one, i.e., one that is composed of the little words **pa + ce + noi + ...,** etc. For it, too, is V-final and free of CC's. Only **bragai** (BRAA-gigh) is a predicate; while it is also V-final, it contains the consonant-pair /br/. Eventually you will be able to translate this sentence as 'Arthur is no longer king'. But right now we are concerned only with its morphology.

The forms of names are discussed in the next section. The morphology of structure words begins with Sec. 2.18, and of predicates, with Sec. 2.34.

A. THE MORPHOLOGY OF NAMES

2.4 The Forms of Names: A name is any sequence of two or more sounds, including irregular ones, in any order except that the last must be a consonant and the whole followed by a pause. The formula for names is therefore (X)+[C/q/x]+.. There is one restriction and one exception to this formula. The restriction will be discussed in the next paragraph. The exception is that there is one very special class of scientific names, the so-called "Linnaean binomials" of biology, whose spellings are fixed by international convention and may not be altered. Occasionally these genus and/or species names end in vowels, e.g., **Escherichia coli.** Occasionally they end in consonants, e.g., **Australopitheicus afarensis.** Thus the formula for Linnaean names is (X)X.. As these exceptional names are irregular beyond the hope of any ordinary morphological redemption, certain extraordinary arrangements must be made to place them in special contexts from which their resolution will be possible in Loglan. These arrangements are described in Sec. 2.13-15. In all intervening sections the word 'name' will refer to non-Linnaean names.

2.5 The Pause Before Vowel-Initial Names: If a name commences with a vowel, then it must also be preceded by a pause. Thus the English names 'Ellen' 'Eileen' and 'Iona' go into Loglan as **Elyn, Aili'n** and **Ai,onas.** So 'Hello, Ellen', 'Come in, Eileen' and 'Give it to Iona' all require pauses in Loglan: /LOI.ELyn/ **Loi Elyn,** /nenKAa.aiLIN/ **Nenkaa, Aili'n** and /DONsula.aiLIN.da/ **Donsu la Ai,onas, da.** (Note the final /s/ added to /aiONa/ to make it C-final; there is more on this move in Sec. 2.11.) If the preceding word is a "name-marker", like **loi** or **la** (see next section), or another name, then the pre-nominal pause is not represented by a comma in text. If the preceding word is anything other than a name-marker or a name, as in **Nenkaa, Aili'n,** then the pre-nominal pause is represented by a comma in text.

The pre-nominal pause is almost always "intervocalic"; see Sec. 2.36 for a discussion of intervocalic pauses.

2.6 The Name-Marker Restriction on Names: The restriction on the (X)+[C/q/x]+. name-making formula arises out of the fact that Loglan names are morphologically so irregular that in order for them to be uniquely resolvable in the Loglan speechstream they must not only be followed by pauses but surrounded by them--as in **Takna, Djan, mi** = 'Talk, John, to me!'--whether they are V-initial or not, unless they are initial in an utterance (**Djan, takna mi** = 'John, talk to me!') or preceded by a "name-marker" (**Takna mi Hoi Djan** = 'Talk to me, O John!'). **Name-markers** like **Hoi** are thus a morphologically privileged class of words. They serve to link C-initial names pauselessly to the rest of the utterance. For example, in the production /TAKnalaDJAN/ (TAAK-naa-laa-JAAN) the name **Djan** is not preceded by a pause. The reason it isn't is because it is preceded by the name operator **la,** which is another member of the class of name-markers that make pausing before C-initial names unnecessary. This last production resolves as **Takna la Djan** and means 'Talk to John!'. It is clear that the name-marker **la** is crucial to its resolution; see Sec. 2.17.

There are just seven simple little words words that have this pause-blanketing privilege, namely I /.i/ **Hoi la loa loi sia siu,** plus any of the compounds formed with

initial /.i-/, e.g., **Ice** = 'And'. Notice that I-words including I itself are themselves preceded by a pause and so must be initial in their breathgroups. But the other six name-markers do not require pausing either before or after they are spoken and so may occur anywhere in a breathgroup. Only **la**, by the way, must be followed by a name; the other name-markers may or may not be. This set of seven simple words and the I-compounds are both all the words and the only words which may precede C-initial names pauselessly.

Restriction is the other side of privilege. Obviously no copy of a name-marker may occur either initially or with a resolvable "prequel" within the first N-2 phonemes of a name. (The last two phonemes of a C-final breathgroup are bound to be all or part of the resident-name, and so do not have to be included in the search for name-markers.) If ***Taknaladja'n**, for example, were offered to us as a name, we would protest that its first part is not a name, but the incomplete phrase **Takna la...**, because that is how we have already resolved it. What we would have to do if it were new to us is first spot the name-marker **la** or a copy of it, and then discover that its prequel, **takna**, resolved. The **prequel** of a copy of a name-marker is that portion of the embedding breathgroup that lies to its left. A prequel **resolves** if the resolver can resolve it completely into regular words, that is, into structure words and/or predicates. Obviously it can do this with /TAKna/.)

So it is clear why this restriction is necessary. Indeed if we wanted to make a name out of the utterance **Takna la Djan**--as we are free to try to do in Loglan--all we have to do is remove that offending sequence /la/. Is **?Taknadj'n** then a name? Yes it is. It meets all the requirements for names including the name-marker restriction. For no copy of any of the sequences /.i... la hoi loa loi sia siu/ occurs within its production as /taknaDJAN/. We might translate such a name into English as 'Talkerjohn'...which opens up some interesting possibilities.

2.7 Working Around the Name-Marker Restriction: The name-marker restriction sounds as if it might have eliminated a large proportion of the objects otherwise usable as names in Loglan. This is apparently not the case. In the many years I have been living with this restriction, I have found only two proscribed sequences which occur in natural names with sufficient frequency to justify a routine response. These are /i/ initially and /la/ either initially or in the context /VlaCX; and both problems are easily dealt with. All other proscribed sequences are exceedingly rare. In fact I remember only two occasions on which a transcription of a natural name had to be rejected because a copy of some name-marker other than /i la/ was resolvably found in it; but the transcriptions were easily modified and I have now forgotten what the problem words were.

The routine solutions I use for modifying /i/- and /la/-containing transcriptions of natural names are as follows. In names like 'Ibañez' 'Isabel' and 'Ibiza'--Spanish is especially rich in /i/-initial names--I replace the /i/ (ee) in the transcription with /ii/ (yee). Thus /iBANieq/ /isaBEL/ and /iBIQa/ are the transcriptions and **Iibanieq** (yee-BAAN-yeth), **Iisabe'l** (yee-saa-BEL) and **Iibiqas** (yee-BEETH-aas) are the Loglan names. /la/-containing names are perhaps a little more common and many come from French. I used to replace the /a/ in /la/ with /e/; but these days I find the /la/ to /ly/ ((laa) to (luh)) as somehow closer to the French. So I now import French names like 'La Fontaine' (laa-fo^{n}-TEN) and 'de la Roche' (duh-laa-ROSH) as Loglan **Lyfonte'n** (luh-fohn-TEN) and **Dylyro'c** (duh-luh-ROHSH).

No doubt other languages and problems will inspire other solutions.

2.8 Derivations of Names: All Loglan names that have been coined so far have been derived from some pre-existing linguistic source. Some sources have been **internal**, from within the Loglan language; others **external**, from outside it. Internal names are discussed in the next section; external ones in Sec. 2.10.

2.9 Internal Names: These are the names that have sources within the language. For example, our family has long had a dog named **Cimr** (SHEEM-rr); her name is derived from the Loglan word for 'summer', which is **cimra**, and certainly fits her disposition. (**Cimra** means 'is the summer of year...'.) So a predicate was converted into a name of a puppy by dropping its final vowel, following it with a pause in speech or a comma in text, and capitalizing its initial letter in text. The result is **Cimr** as in **Cimr, kamla** (SHEEM-rr . KAAM-laa), or, for that matter, in 'Cimr, come!' Another example. We once had a cat with the uninspired name **Gro'katm**, which means Big Cat (from **groda katma**). We pronounced it (GROH-kaat-mm); but we probably shouldn't have. We were deriving the name from the complex predicate **grokatma** (groh-KAAT-maa) in which (KAAT) is the stressed syllable; so the name could have been regularly stressed as **Grokatm** (groh-KAAT-mm). Still, the apostrophe in the written form **Gro'katm** would have made it clear in text--had it ever gotten into text--that we were stressing it irregularly...as one is free to do with names.

Any predicate may be used to yield a name in this way. Thus, if one wants to imitate the Caribbean custom of saying 'Man!' vocatively, Loglan **Mren**, used in such expressions as **Mren, ea mu safgoi** /MREN.eAmuSAFgoi/ = 'Man, let's sail!', does the job nicely. (**Safgoi** is derived from **salfa godzi**, which means 'sail-go'.) **Fum** is an equally useful vocative, coming as it does from **fumna** = 'is a woman'. Some other internally derived names come from structure words. These usually require that a consonant be added to the V-final word to make the name, and by convention the consonant added to make internal names is /n/. Thus **Tun** comes from **tu**, which, when used as a vocative, has the sense of 'You!' or 'Hey, you!'.

2.10 External Names: Externally derived names are the closest possible imitations in Loglan of either (a) the **sounds** or (b) the **appearances** of certain natural language names which are to be imported into Loglan. It is generally not possible to do both.

2.11 Auditorily-Modeled External Names: When the things named are persons or places, it is conventional in Loglan to imitate the sounds of the natural name as closely as this may be done in Loglan phonetics. The preferred auditory model is the way the name in question is, or would have been, pronounced by the person named, or by the people who live in the place named. If that natural pronunciation does not end in a consonant, then it is conventional to add the phoneme /s/ to external names. Thus if Peter is an Englishman, his own pronunciation of 'Peter' will sound something very like (PEET-uh) in Loglan phonetics; whence **Pitys** is his Loglan name. But if Peter is an American, we will hear (PEED-rr) when he tells us who he is, so no /s/ will be required to complete his Loglan name. **Pidr** and **Pitys** are two very different-sounding names; and in Loglan they are spelled differently as well.

The procedure is similar for names of places and famous personages. 'Bach' is **Bax**

and 'Berlin' is **Berli'n** in Loglan because (bahkh) is a good copy of the German name of that German composer and (behr-LEEN) is an almost perfect reproduction of how Berliners speak the name of their city. 'London', on the other hand, is **Lyndn** because (LUHN-dnn) is another very good copy of the sounds one hears on the lips of Londoners despite the unsettling departure from English spelling that is caused by writing it phonetically. (English, as everyone knows, is an unphonetically spelled language.) 'Paris' suffers hardly any visual change because in being phonetically rewritten as **?Pari'** (paa-REE) it becomes clear to the word-maker that it requires the conventional addition of final /s/; thus **Pari's** is Loglan for 'Paris', an adventitious similitude. 'Roma' is **Romas** because again the local pronunciation (ROH-maa) is V-final and again /s/ is conventionally added to the original, this time without an adventitious matching of the local spelling.

In general, when borrowing place and person names, loglanists try to make the best imitations possible with their limited phoneme kit unless the result is thought to be too difficult for loglanists to use on a world-wide basis. Thus, should German 'München' (Munich) be exactly reproduced as **Mwnxen** (MEUN-khen) (as it can be by using the irregular phonemes /w/ = (eu) and /x/ = (kh) which happen to be German)? Or should we adopt the more forgiving **Muncen** (MOON-shen), which although quite close to the original is still pronounceable by that large population of potential loglanists whose tongues could not reproduce the German sounds exactly? This is not a morphological question but a broader linguistic one on which The Institute has no policy at present. Until it is settled by an international congress of loglanists, current loglanists are free to adopt whatever degree of imitation they can manage. Thus either **Mwnxen** or **Muncen** but certainly not ***Miunek.**

2.12 Visually-Modeled External Names: When the name to be imported is a scientific or scholarly one, and a certain method of spelling it is in wide international use, then the model to be most closely copied is the visual appearance of the internationally shared portion of that name as spelled (or respelled) in the Latin alphabet. Take the names of the planets in the five major European languages:

English	Russian	Spanish	French	German
Mercury	Merkurie	Mercurio	Mercure	Merkur
Venus	Venera	Venus	Venus	Venus
Mars	Mars	Marte	Mars	Mars
Jupiter	Yupiter	Jupiter	Jupiter	Jupiter
Saturn	Saturn	Saturno	Saturne	Saturn
Neptune	Neptun	Neptuno	Neptune	Neptun
Uranus	Uran	Urano	Uranie	Uranus

From the appearances of these words, the choices all seem obvious. The only question would appear to be: Should we add /s/ or subtract the few final vowels when there are any? Clearly we should do the latter. The greatest variation in these natural names is precisely in these final vowels when they exist at all. This decision gives us either **?Mercur** or **?Merkur** for 'Mercury'; **Venus** for 'Venus'; **Mars** for 'Mars'; **Jupiter** for

'Jupiter'; **Saturn** for 'Saturn'; **Neptun** for 'Neptune'; and **?Uran** or perhaps **?Uranus** for 'Uranus'. Only the choices between **?Mercur** and **?Merkur** and between **?Uran** and **?Uranus** remain. **?Mercur** would appear to imitate the slightly more widespread spelling. On the other hand, the pronunciation of [c] in the [Mercur-] sequence in all the languages that use it is Loglan /k/. This consideration tips the balance, in my opinion, in favor of **Merkur**. As for Uranus, **Uran** is clearly the best choice; it is the internationally-shared portion of these five written names, and its sequelae are extremely variable.

But now let us consider the pronunciation of these visually good copies of these low-variance portions of these six international words. The copies are so good, in fact, that in the astronomic or myhthological contexts in which they are most likely to appear in print, they will surely remind any scholarly reader of what they mean. But how do they sound? Quite Loglandical, it turns out: (MEHR-koor), (VEN-oos)--not (VEEN-uhs) or (VEIGHN-oos)--(mars)--that's an /s/, not a /z/--(zhoo-PEET-ehr), (SAAT-oorn), (NEP-toon), and (OOR-aan). Only one, **Neptun**, sounds like the English word. But that is just as it should be. These words are international words. They have, like all the tools of shared scholarship, only international looks, not international sounds. If we in the Loglan community pronounce them in our special way, we are only doing what the rest of the international community of scholars has been doing all along.

2.13 The Linnaean Polynomials: The Linnaean "binomials"--actually, polynomials because some of them have three parts--are a huge body of zoological and botanical terminology--probably numbering well into the early millions by this time--that have been made standard in spelling by international covenant among biological scientists. The covenant is very simple. Provided a proposed new Linnaean name is unique, then however it is spelled on its first appearance in print (including any adventitious errors) is the way it will then be spelled by all who use it afterwards. Even Japanese and Chinese scholars, including those who use the Latin alphabet in no other way, use Latin letters and Linnaean spellings to represent species and/or genera in the text of their own scientific writing...even if that writing is intended only for domestic consumption.

This stern system is called "Linnaean" after the great Swedish naturalist Carl von Linne who proposed this remarkably enduring system of biological nomenclature several centuries ago. Loglan has no choice but to follow it exactly. It is clear that the names of **Homo sapiens neanderthalensis**, **Escherichia coli** and **Australopithecus afarensis** must be at least written identically in all languages...even if they are pronounced in a great variety of ways. Phonological diversity doesn't matter; what matters is that readers will continue to be presented with the same words. So here the job of The Institute is to find a way of pronouncing these letters that is (1) consistent with the phonology of Loglan, (2) modeled on some natural language way of pronouncing these polynomials that is imitable by speakers of the widest possible group of other languages, and (3) does not--as far as can be foretold by examining reasonably large samples--render any two Linnaean words identically in speech. Once that is done, then a way must be devised for transforming the binomials spoken by loglanists into Loglan text that will not cause us to deviate from the audio-visual isomorphism of the language.

As far as pronunciation is concerned, the habits of Romance-speaking scholars, especially Italian and Spanish ones, commend themselves on all three counts. Except for the pronunciation of [C c], Romance pronunciation is consistent with Loglan phonology--which is a phonological close relative--and it employs, like Loglan, a minimum set of widely used phonemes. Moreover, the Spanish and Italian communities

of biologists have been pronouncing Linnaean words for several centuries; and it is unlikely that the pronunciation habits they have developed for that purpose produce homonyms with detectable frequency.

First, we will estimate this Romance-based pronunciation pattern; then in Sec. 14, we will consider how isomorphism in the neighborhood of Linnaean words may best be preserved.

2.14 Pronunciation of the Linnaean Polynomials: Here is our current estimate of how an Italian or a Spanish biologist, with some knowledge of how the French, German, English or Slavic proper names that are often celebrated in the polynomials are "really pronounced", as well as some feeling for the spelling peculiarities of Loglan, would try to regularize the pronunciation of the Linnaean vocabulary:-

We suppose that such a scholar will pronounce every letter and letter combination as in standard Loglan except these:-

[ae]	(pronounced as Loglan)	/e/	(eh)
[au]		/o/	(oh)
[aw]		/o/	(aw)
[ay]		/ei/	(ay)
[c]	before [a o u] or C not h	/k/	
[ch]	initially & finally	/tc/	(ch)
[ch]	medially except in [sch]	/k/	
[ea]	except final & in [eau]	/i/	(ee)
[eau]		/o/	(oh)
[ee]		/i/	(ee)
[eigh]		/e/	(eigh)
[eu]		/oi/	(oy)
[ew]		/u/	(oo)
[igh]		/ai/	(igh)
[j]		/dj/	(j/dg)
[h]	before any C	/hy/	(huh)
[lh]		/l/	
[oe]		/o/	(oh)
[ou]		/u/	(oo)
[ow]		/ao/	(ow)
[ph]		/f/	
[p]	before [n s t]	/py/	(puh)
[sh]		/c/	(sh)
[sch]		/c/	(sh)
[q]		/k/	
[rh]		/r/	
[th]		/t/	
[w]	before V	/u/	(w)
[x]	initially	/z/	
[x]	non-initially	/ks/	
[y]		/i/	(ee)

Note that, in deference to our Loglan habits, our standard-setting Romance-speaking scholar will pronounce [C c] before [e i y] as Loglan /c/, that is, as written, and as /k/

elsewhere. Thus 'Cephalopoda' is to be pronounced (sheh-faa-loh-POHD-aa), while 'Coelenterata' is pronounced (koh-len- tehr-AAT-aa). This follows half the European custom of pronouncing [c] before "strong vowels" ([a o u]) like Loglan /k/, and before "weak vowels" ([e i y]) like /s/. Instead of /s/ we will ask our standard- setter to use our "native" Loglan /c/.

Double consonants, that is, pairs of identical consonant letters, will be pronounced as if they were a single instance of that consonant.

2.15 Writing Linnaean Names: Once Linnaean names are being pronounced in a standard way by loglanists, then the problem arises of how to furnish Loglan readers with clear indications of which names are Linnaean in a given text. Since some Linnaean names are vowel-final (**Homo**) and some are not (**sapiens**), their endings cannot be our clue; and while we are obliged like the rest of the scientific community to italicize, bold or underline Linnaean names in text, even these textual indications are not unique and are in any case difficult to represent unequivocally in speech. So in the interests of preserving isomorphism, a special Linnaean name operator is required. Instead of **la** we will use **laa** (LAA-aa) for announcing Linnaean names. We will further specify that **laa** always be followed by a pause in speech and that Linnaean names never be used without this operator, that is, as vocatives. (So you may not address your visitor as ***Hoi Canis lupus** even though da is one.)

Moreover, the variable number of terms in these polynomials--sometimes monomials (**Australopithecus**), most freqently binomials (**Escherichia coli** or **E coli**), but sometimes trinomials (**Homo sapiens neanderthalensis**)--would mean that we could not tell when a V-final breathgroup following a Linnaean term was simply another Linnaean term or a string of regular words. The threat of this ambiguity forces us to treat all Linnaean names as monomials in speech, looking to a separate algorithm to partition these transitional monsters, e.g., **Homosapiensneandertalensis** (hoh-moh-saa-PEE-ens-neigh-aan-dehr-taa-LEN-sees), **Kanislupus** (kaa-nees-LOOP-oos), back into their proper, and properly-spelled, polynomial parts. Because of their length we anticipate that some of these transitional monomials will be multiply stressed; but it is doubtful that the existence of these occasional multiple stresses will be more than adventitiously useful in partitioning.

The algorithm that will complete the isomorphism has not been written. And when it is written it will probably involve repeated lookups in extremely long tables. But the principle of that algorithm is clear. Loglan can continue to be isomorphic, even in the region of the Linnaean polynomials, if a machine can be programmed to imitate the performance of the microbiologist (or the icthyologist, or the botanist, or whatever, who hear Linnaeans from their specialties) who can listen to /eceRIKiaKOLi/ and know that da has heard a Linnaean name that da can confidently type as [**Escherichia coli**], including the bolding that is now conventional.

We know that such programs can be written; it is just a question of waiting for some competent loglanist with a biological bent to find the time.

2.16 The Post-Nominal Pause: Many loglanists have misunderstood the requirement that there be a pause after names to mean that, in order for Loglan speech to be understood at all, these "obligatory pauses" must be carefully produced every time. Not so. Sloppy Loglan has about as good a chance of being correctly understood--at least by a sympathetic human listener--as sloppy speech in any natural tongue...which is

to say that the chances are pretty good. The difference is that non-sloppy Loglan speech can be impeccably understood by any properly instructed listener...even a machine.

2.17 Resolving Names: The resolution of a Linnaean name from the speechstream is trivial. One comes upon the Linnaean operator **laa** followed by its pause; one scans right for the next pause or the end of the utterance; all that lies between the two pauses, or between the pause and silence, will be the Linnaean name. Partitioning the monomial thus generated into its proper Linnaean parts, and respelling those parts in ways that recapture the original spellings exactly, are non-trivial steps but beyond the scope of this book.

The resolution of non-Linnaean names is also non-trivial but quite easy to follow. Before the resolver can resolve names from the speechstream it must know how to resolve regular words from it; for a crucial step in the resolution of most names is the resolver's attempt to resolve as regular words the prequel of some copy of a name-marker that has been found in it; see Sec. 2.6. Thus the algorithm described in this section presupposes the existence of algorithms like those that will be described later for little words and predicates; Secs. 2.33 and 2.59-60. However, our discussion of the name-resolving algorithm can be followed without reference to those later algorithms. All that is required is the suspension of disbelief that they can be written.

Given that a regular-word resolver exists, then, the first step in the resolution of a name is to:

1. Find the first (or next) breathgroup in the utterance that is C-final. If there are none, then there are no names. (Being followed by obligate pauses, all non-Linnaean names are either right portions of their breathgroups or occupy them entirely. So all and only consonant-final breathgroups have resident names.) Suppose we locate a C-final breathgroup.

2. Search the first N-2 phonemes of the breathgroup for all copies of the sequences /.i la hoi loa loi sia siu/. If none are found, the entire breathgroup is a name and may be so resolved. (The last 2 phonemes in the breathgroup do not have to be searched; they will always be either part or all of the resident name.)

3. If one or more copies of the specified sequences are found, select the rightmost and discover whether its prequel will resolve into regular words. If its prequel does resolve, the name commences just to the right of that copy and stretches to the end of the breathgroup and may be so resolved. If the prequel of some copy does not resolve, select the next rightmost copy if any remain and repeat this step until no copies remain.

4. If no copy with a prequel that resolves into regular words can be found, then the entire C-final breathgroup is a name. Return to step (1) for further breathgroups until all C-final breathgroups have been processed in this way.

B. THE MORPHOLOGY OF STRUCTURE WORDS

2.18 The Functions of Structure Words: The shortest words of any language are its logical, numerical or grammatical words. These are words like 'or' 'of' 'is' 'a' 'if' 'one' 'two' 'the' and endings like '-s' '-ed' '-ing' and '-hood' in English. . In Loglan short morphs of this kind include all the grammatical particles (tense words, articles, etc.), the connectives (logical and causal), all the prepositions that are not predicates, the case-markers, the pronouns, the other variables, and all the number- and letter-words.

Clearly, structure words are the words that shape the **structure** of an utterance, and into which the less frequently-used and typically longer **content** words, the predicates and names, are fitted as pictures into frames. Thus, **Le** _____ **pa** _____ is a sentence-frame. Its nature is completely determined by two structure words. But any two predicates we care to choose can complete the picture so-framed. For example, let us use **mrenu** and **fumna.** If we drop these two predicates into the empty places in the frame in both possible orders, we will make two sentences out of them: **Le mrenu pa fumna** = 'The man was a woman' and **Le fumna pa mrenu** = 'The woman was a man'. If there are 10,000 predicates in the language, then there are 100,000,000 ways of filling this one frame. Each structural frame is therefore a set of possible sentences.

This is the function of structure words: to build the structural frames which the content words then fill out, thus creating the utterances of the language.

2.19 The Four Little-Word Forms: All **simple structure words** in Loglan are short; the longest have only three letters. Since the shortest predicates are four letters long, there is a complete visual and audible separation of these two kinds of regular words in Loglan. Therefore we will occasionally call the simple structure words **little words,** for in Loglan they are genuinely little. It is the **compound little words** described in the next section that attain substantial lengths.

The four forms of little words are .V VV CV and CVV. Using a typical member to represent each form-class, we can call them the **A**-form words, the **Ia**-form words, the **Da**-form words, and the **Tai**-form words.

The complete set of **A**-form words are **a e i o u.** In ordinary contexts, these are the five simple logical connectives 'or' 'and' 'if and only if' and 'whether' etc., although they are also used as letter words in spelling contexts; see Sec. 2.25. These five tiny words, as well as any compounds which commence with them, must always be preceded by pauses in speech and commas in text. For example, in **Da, a de pa kamla** /DA.aDEpaKAMla/ ('X and Y came') the pause-comma is obligatory.

There are 25 **Ia**-form words and all of them are attitude indicators. Examples are **ia** and **io** ('Certainly' and 'Probably'). No pause need accompany these **Ia**-form words or their compounds: **Da ia, e de io pa kamla** /daIA.edeIOpaKAMla/ (daa-YAA . eh-deh-YOH-paa-KAAM-laa) = 'X certainly, and Y probably came'.

Next are the 85 **Da**-form words formed of the 5 primary vowels combined with the 17 regular consonants. All possible **Da**-form words have currently assigned meanings. Indeed, the CV form is perhaps the hardest-working morphological form in the language, in the sense that more CV morphs occur in Loglan utterances than morphs of any other form. The best examples of **Da**-words are **da** itself and its four companions, **de di do**

du. These are often well-translated into a sort of "mathematized" English by the gender-less, number-less, case-less mathematical variables 'X' 'Y' 'Z' 'W' 'Q'. For example, **Da kepti de di do du** means 'X is a ticket to Y from Z on carrier W for price Q'.

Finally, there are the more numerous but less frequently used **Tai**-form words. There is morphological space for more than 400 of these CVV-form morphs, and only about half of them have been assigned; so we still have plenty of room for growth.

> [It is a widely believed myth among loglanists that we are "running out of CVV space." This is not true. The last 12 years of active lexical development have added only a few dozen **Tai**-form words to the 148 words of this form that had meanings in 1975. More than a few dozen empty places, namely 85, have been added morphologically to the CVV word-space since that time. Five new phonemes /h y x q w/ were added to the language--even though four of them, /y x q w/, added only letter-words to the CVV space, and then promptly occupied it--and the previously unused vowel-pairs /aa ee oo/ have been discovered to have uses and so have augmented the rows of the CVV table. Since the number of **Tai**-form words that have actually been added to the language, including the new letter-words, is far less than 85, we have even more open CVV-space now than we had in 1975.]

The best examples of **Tai**-words are the letter-words **Bai Cai Dai** etc. [B C D etc.] of which **Tai** [T] itself is one; but many of the prepositions and adverbs of the language are also **Tai**-form. For example, **dio** is a "case tag" that means 'to/toward' and **sui** is a "discursive" that means 'also'.

As we shall see presently, the 3 irregular sounds /q w x/ and /y/ are fully represented morphologically only among the letter-words. In fact, CVV is the only little-word form in which /q w x y/ may occur. All other little words are made from the 5 primary vowels and the 17 regular consonants.

2.20 Compound Little Words: A compound little word is a string of simple little words concatenated pauselessly in speech and printed without spaces in text. Thus, **pacenoina** is a compound little word, or simply a **compound.** It is composed of **pa + ce + noi + na.** Most combinations of little words are permitted in compounds. One exception is that VV-form words may be compounded only with each other. For example, **uaui** expresses a satisfied kind of happiness and **uiua** expresses a happy kind of satisfaction, and both are compound attitudinals. Compound attitudinals may be of any length; but they are the only kind of compound in which VV words may occur.

There is one more restriction. While V, CV and CVV words may be mixed together in a compound, two orders are proscribed: V words may not follow other V words (the result would look like a VV word), and they may not follow CV words (which result would simulate a CVV word). So V words may in fact only follow CVV words, which they do only rarely and then only when the **Tai**-word is of Cvv-form, that is, only when it contains a monosyllabic vv-pair like **Tai** itself. When V words do follow Cvv words in a compound, they lose their leading pause. That pause is retained when a .V word is used initially in a compound. Thus .V + CVV puts the obligatory pause in front of the compound where it belongs. For example, **anoi** = 'if' is such a word, and the leading pause is preserved in use: **Da, anoi de** /DA.anoiDE/ = 'X if Y'. *Cvv + .V, in contrast, would put the pause in the middle of the word if the pause were retained. So it is not retained. (Words may have no medial pauses, of course; see Sec. 2.28.) An

example of a Cvv + V compound is **MaiA** (MIGH-aa), the acronymic word for [MA], the US Postal Service's acronym for 'Massachusettes'. Note that /MAIa/ (MIGH-aa) is audibly distinct from /MAia/ (MAA-yah). This phonological distinction will be employed in the resolution of little words; Sec. 2.33.

There are also some "false compounds" that we must look at. Derivations of the CV + V kind do exist semantically even though morphologically they are proscribed. For example, the CVV word **noa** has a meaning which is derivable from **no** followed by **a** ('not and/or' or 'only if'). But the resulting **noa** is not a compound. It is a simple CVV word chosen with that semantic derivation in mind. This kind of derivation applies to the **Tai**-form letter-words as well. Thus **Tai** itself is semantically derived from the letteral [T] plus the suffix [ai]. But **Tai** is nevertheless a simple CVV word, and so is morphologically not subject to further resolution.

2.21 Letter-Words: These are a kind of structure word which have a very special morphology in Loglan. By a **letter-word** is meant a word like English 'em', 'eff', or 'dee' by which letters are spoken or read aloud from text in a given language. That they are words in the spoken forms of languages which have written forms, and therefore have characters of some kind--and of course, not all languages do--is clear. We can say 'There are nineteen effs on this page' in spoken English. But words like 'eff' seldom appear in English text. Instead, in writing such a sentence we would probably use the letteral [f] and type [There are nineteen f's on this page.]. In Loglan, letter-words appear as frequently in the written language as in the spoken one. There are also a great many Loglan letter-words, since there are separate words for the upper- and lower-case versions of each letter in the Greek and Latin alphabets. That makes 100 letter-words in all, as there are 26 letters in the Latin alphabet and 24 in the Greek one.

Each Loglan letter-word is formed by combining the Loglan phoneme associated with that letter with a suffix. If the letter is Latin, the association is automatic. The phoneme associated with a given letter is the sound that letter is given in reading Loglan text aloud. If the letter is Greek, however, some of the associations between characters and sounds are obvious and some are arbitrary. So the entire list of associations between Greek characters and Loglan sounds will be given presently.

2.22 Suffixes for the 52 Latin Letter-Words: Some of these words are **Tai**-form; some of them are **Ama**-form. The four suffixes required to generate all 52 Latin letter-words are as follows:

For the 7	lower case	Latin vowels,	add **-si;**	thus **asi, esi,..., ysi.**
	upper		add **-ma;**	thus **Ama, Ema,..., Yma.**
19	lower	consonants,	add **-ei;**	thus **bei, cei,..., zei.**
	upper		add **-ai;**	thus **Bai, Cai,..., Zai.**

In addition, there are the 7 single-letter abbreviations of the vowel letter-words provided by the 7 vowels **a e i o u w y** used as one-letter words. These, too, are letter-words. But unlike the **Ama-** and **asi**-form vowel-words, to which they are alternatives (allomorphs), the single-letter letter-words are ambiguous with respect to both language and case. As we shall see later, the single phoneme /a/ may stand for the upper case Latin letteral **A**, for the lower case Latin **a**, for the upper case Greek alpha (we cannot

display the Greek letterals on our font, so will be content to name them), or for the lower case Greek alpha. Which letteral the vowel /a/ is representing in any given case will depend entirely on the context in which we find it; see Sec. 2.24.

The reason there is a single-letter abbreviation for each of the Latin vowel-words is that, in many contexts in which the Latin letter-words are used, the preferred vowel-word is the vowel itself. Thus in spelling the Loglan word **ba** we wish to name the characters; so we say say **bei, a** (bay . aa). We could say **bei asi** (bay . AA-see) if we chose, thus specifying the lower case Latin [a]; but that much univocality is not required in the spelling context and is seldom used. See Sec. 2.25 for more on spelling practice. In fact, the 7 single-letter vowel-words, **a, e, i,** etc., are used either in or as letter-words wherever the loss of case and language information is unimportant; for example, in making acronymic words; see Sec. 2.29.

Consonants cannot, of course, be spoken alone. Therefore they always require a vocalic suffix.

2.23 Suffixes for the 48 Greek Letter-Words: These words are also of either **Tai**-form or **Ama**-form, but of course four different suffixes are used. The suffixes required to generate the 48 Greek letter-words are as follows:-

For the 6	lower case Greek vowels,	add **-fi;**	thus **afi,** etc.
	upper	add **-mo;**	thus **Amo,** etc.
18	lower consonants,	add **-eo;**	thus **beo,** etc.
	upper	add **-ao;**	thus **Bao,** etc.

Two phonemes in the Loglan phoneme set have no corresponding letters in the Greek alphabet. These are /c/ and /w/, (sh) and (eu). The remaining 24 Loglan phonemes have been tentatively assigned to the 24 letters of the Greek alphabet as follows. There are four with arbitrary associations and these are marked with a pound-sign [#]:

a	alpha	h#	eta	n	nu	t	tau
b	beta	i	iota	o	omega	u	upsilon
d	delta	j#	xi (ksi)	p	pi	v#	psi
e	epsilon	k	kappa	q	theta	x	chi
f	phi	l	lambda	r	rho	y#	omicron
g	gamma	m	mu	s	sigma	z	zeta

As I say, these assignments are tentative. The Institute would be pleased to consider any proposal based on a better understanding of Greek phonemics than this one displays.

To use these tables to build a Greek letter-word, proceed as follows.

(1) Suppose we want the word for lower case Greek gamma.
(2) The suffix for l.c. Greek consonants is **-eo.**
(3) Gamma is associated with the Loglan phoneme /g/.
(4) So the required letter-word is **g + eo = geo,** a **Tai**-form word.

2.24 Uses of Letter-Words: Letter-words are currently being used in five contexts: (1) In spelling, see next section. (2) In making acronymic words like **CaiIzA** (shai-EEZ-aa) for 'CIA' (see-igh-EIGH), see Sec. 2.29. (3) In forming dimensioned numbers

like **nenimei** (neh-NEE-may) [10m] for '10 meters', see Sec. 2.31. (4) As letter-variables both in mathematics (**toXai** [2X]) and in ordinary discourse (**Bai groda Cai** = 'B is bigger than C', which is often abbreviated in text to [B groda C]). (5) To form scientific predicates, for example, **geoykreni** (geigh-oh-uh-KREH-nee) for 'gamma-ray', which is made from the letter-word for lower case gamma, **geo** (GEIGH-oh), the hyphen /y/ (uh) (see Secs. 2.48 and 2.55), and the predicate **kreni** (KREH-nee), which means 'ray'.

2.25 Spelling Aloud: To spell a word aloud in Loglan, one uses **Tai**-words for the consonants and either **A**-words or **Ama** type words for the vowels at the speller's option. Normally da will use **A**-words for the vowel-letters. But if capitalization is to be reported, or there is any other source of confusion in the context, da may choose to use **Ama** words for greater explicitness. Thus the string of utterances (for so the grammar will perceive it) **Tai A I** /tai.a.i/ (tigh . aa . ee) will be taken by the Loglan auditor to spell the word [Tai]. In English, we would say 'Capital tee. Eigh. Eye.' More explicitly da might wish to say **Liu Tai nu leasri li, Tai A I lu** = 'The word 'Tai' is spelled (character-written) 'Tai. A. I.'.' A guide for this string is (lee-oo-TIGH-noo-leigh-AAS-ree-lee . tigh . aa . ee). There is more on spelling in the grammar.

2.26 Little Word Predicates: There are three series of words in Loglan that considered morphologically are compound little words but which are treated by the grammar as predicates. These are (1) **mathematical predicates**, (2) **acronymic predicates**, and (3) **identity predicates.** Identity predicates are **bi** and its analogs and compounds (see Lexeme BI in the Lexicon) and require no special morphological treatment. The other two series of little word predicates do require special treatment and will be discussed in the next section and in Sec. 2.29.

2.27 Mathematical Predicates: There are the two series of these words: the **cardinals** and the **ordinals.** They are generated by attaching either the cardinal suffix /-ra/ or the ordinal suffix /-ri/ to any number word or other quantifier; see Lexeme PREDA for the complete list of non-numerical quantifiers. Examples are **tora** = 'is a dyad/a twosome' and **rari** = 'is the "all-th" or final member of sequence...'. Like other predicates, mathematical ones are stressed penultimately. So they must be separated from any preceding number-word or quantifier by a pause in speech or a comma in text. (Such commas are not strictly necessary in text, but it is considered good writing style to use them because they contribute to the isomorphism of the language.) Thus **Kambei leva fe, fefera galno veslo mi** /KAMbeilevaFE.feFEraGALnoVESlomi/ = 'Bring those five fifty-five gallon containers to me' must be partitioned into (at least) two breathgroups at the juncture between the quantifier and the cardinal predicate to prevent /KAMbeilevafefeFEraGALnoVESlomi/ from being heard. The pauseless production would mean 'Bring that five-hundred-and-fifty-five gallon container to me'.

> [This stress rule is new since 1983. It was decided to bring the stress in mathematical predicates--once initially stressed--into line with that of all other predicates. The cost of this rectifying move is phonologically a modest one: the occasional use of a quite naturally-occurring pause. Morphologically it eliminates an exception. It allows us to say that <u>all</u> predicates are stressed penultimately.]

2.28 The "No Pausing Inside Words" Rule: The preceding resolution demonstrates the necessity of not pausing inside a word...especially not inside a compound one, which will often have some other resolution if the speaker does inadvertently pause. This is because no Loglan word legitimately contains a pause and so the resolver makes good use of whatever pauses it hears. (This means that the tiny stops that do occur acoustically in some vowel disyllables, e.g. in /a,o/, will be measured by the machine--and by the human auditor, for that matter, who is usually unaware of them--as "too brief to be a pause".) New loglanists frequently object to this rule: 'But sometimes I don't know the word, and I have to hesitate until I remember how it goes!' That is true; and human auditors--your teacher, for example--will understand this...and probably you. But the **machine** will not understand you until your Loglan speech becomes "fluent". That, in fact, is what any increase in fluency largely is. It is the elimination of just these morphologically unnecessary (and sometimes downright misleading) pauses from the hesitant speech with which you and every other learner will inevitably begin.

2.29 Acronymic Predicates: Morphologically, acronymic words are compound little words; but grammatically they are predicates. They are one of several classes of little word predicates (see Sec. 23) which have this slightly misleading morphology.

But what are "acronymic words"? Just as a letteral, let's say [T], is a visual abbreviation of its letter-word **Tai**, so a Loglan acronym is a visual abbreviation of its acronymic word. Thus wherever [CCC] occurs in Loglan text it is simply an optional, written abbreviation of the word **CaiCaiCai**; and both are pronounced (shigh-SHIGH-shigh), for, like all predicates, acronyms are stressed penultimately. This is not quite like the handling of acronyms in English. The English acronym [CCC], for example, is a representation in writing of the "spelling-form" 'See-see-see'; but this form never occurs in written English. So a closer parallel with Loglan acronyms is the use of compound numerals in both languages. Thus, [123] is a shorthand written notation for the spoken expression 'one-hundred-twenty-three' in English just as [123] is shorthand for **netote** in Loglan. Both "longhand" forms may occur in writing as well as in speech. Just so, [CCC] is shorthand for **CaiCaiCai**; and **CaiCaiCai** is not only the "spelling aloud" of [CCC] but it is a written word as well.

Conventionally, the acronyms of Loglan are restricted to Latin letterals, both upper and lower case. The variety of internationally-used acronyms that can be "spelled out" as acronymic words in Loglan includes not only common sequences of Latin upper-case letterals, like [USA UN DNA], but also alphanumeric sequences like [U234] and chemical formulas like [H2SO4]. Also, chemical symbols containing both lower- and upper-case letterals, such as [Fe] (Iron) and [As] (Arsenic), are also uniquely spellable as Loglan acronymic words once certain decoding conventions are taken into account. But first let us look at the rules for turning such acronyms into Loglan words. These "spelling-out" rules are as follows:-

1. All consonant-letterals in an acronym are represented in the acronymic word by their full 3-letter words. Thus **peicei** is the word of which [pc] is the written abbreviation; and both expressions are pronounced /PEIcei/ (PAY-shay).

2. With a very few exceptions to be noted later, each Latin vowel-letteral in an acronym is represented by that single vowel phoneme in the acronymic word. Thus **DaiNaiA** = /dai + nai + a/ is the reading-aloud, or spelling out, of **DNA**; and both expressions are pronounced /daiNAIa/ (digh-NIGH-aa).

3. If a pair of Latin vowel-letterals are adjacent in an acronym, the vowel phonemes by which they are read aloud are hyphenated with /z/ in the acronymic word. Thus **CaiIzA** = /cai + i + z + a/ is [CIA] read aloud. Both are pronounced /caiIZa/ (shigh-EEZ-aa). Similarly, **AzAzA** (aa-ZAA-zaa) is [AAA] read aloud.

4. If there is a 2-letter element symbol in a chemical acronym which, like [Ca] (Calcium), is composed of an upper-case consonant-letteral plus a lower-case Latin vowel-letteral, the two are hyphenated with /z/ in the acronymic word. Thus, [Ca] itself is read as **Caiza** (SHIGH-zaa) while [CaCO3] (calcium carbonate) is read as **CaizaCaiOte** (shigh-zaa-shigh-OH-teh). Note that [O] is not hyphenated to the second [C]. Thus [O] is not a lower-case appendage of that [C].

5. To read an acronym containing Greek vowel-letterals as a Loglan word, those vowel letters must be read aloud by their full three-letter words, e.g, /AMo/ or /AFi/. Any Greek consonant-letter in an acronym must be hyphenated with /z/ to any immediately following vowel of either nationality.

6. All non-initial numerals in an acronym are pronounced as number-words in the acronymic word. Thus **Utotefo** = /u + to + te + fo/ is the word of which [U234] is the abbreviation. Both are pronounced /utoTEfo/ (oo-toh-TEH-foh).

7. Acronyms with initial numerals--rare forms at best--are not allowed to be transformed into Loglan acronymic predicates. If they were, they would be taken for dimensioned numbers, e.g., **temei** = [3m] (TEH-may) 'three meters'.

8. Dimensioned numbers may have acronyms as their right-hand parts. Thus [100USD] might be the written form for 'one hundred US dollars', in which [USD] is an acronym. The whole expression, then, would be pronounced /nema,uSAIdai/ (neh-maa-oo-SIGH-digh) in which the resolver would detect the /u/ as part of the acronym [USD] and therefore not pair it with the preceding /a/. Written out, the compound word for [100USD] is **nemaUSaiDai.**

9. Acronymic words whose acronyms would imitate existing or even possible Loglan words are not permitted. This ***peia** is not permitted because its acronym is [pa] which imitates **pa**; see Sec. 28 for the acronym recovery rules. In speech /PEIa/ and /pa/ would be quite distinct, but in written text the acronym [pa] would be indistinguishable from the word [pa].

2.30 Pause and Stress Around Acronymic Words and Letter-Words: Stress is always penultimate in acronymic predicates, that is, stress falls on the syllable that is second from the last, as indeed it does in all predicate words.

Given the penultimate stress rule, Rule 5 in the preceding section requires that the juncture between single-vowel connectives and a quantifier, as in **U totefo le mrenu** ('Whether 234 of the men'), be protected by either a pause or by final stress on the quantifier: thus either /u.totefoleMREnu/ or /utoteFOleMREnu/ would be morphologically distinct from the production with the penultimately stressed acronymic word, namely /utoTEfoleMREnu/ ('Be U234 (to/at/about) the men(?|)' whatever that would mean). My prediction is that usage will follow the second, more economical route. But let us see.

The fact that acronymic words, being predicates, are always penultimately stressed may also be used to force the resolution of a string of separate letter-words (which otherwise might imitate an acronym) by stressing their final member. Thus, in /VEDmabaicaiDAI/ the sequence /baicaiDAI/ cannot resolve as the acronym **BaiCaiDai** [BCD], for that would have to be penultimately stressed: /baiCAIdai/. Thus the resolver can write **Vedma Bai Cai Dai** = 'Sell B to C for D', or even [Vedma B C D], from the production /VEDmabaicaiDAI/ without benefit of pauses. Pauses may of course be used to separate acronymic words from each other and from the number- and/or letter-words they might otherwise absorb, or optionally (never obligatorily) to separate letter-words from one another. Thus **Bai, DaiNaiA Cai** = 'B is the DNA of C (that is, part of C's genome)' may be univocally produced with only that one pause: /BAI.daiNAIacai/ or, just as effectively, /BAI.daiNAIaCAI/. However, if we drop the pause, the resulting production /BAIdaiNAIacai/ resolves as **BaiDai NaiA Cai** ('Be a BD type of NA to/of/about, etc. C'), and if at the same time the stress is dropped from /BAI/, then the resulting production, /baidaiNAIacai/ resolves as **BaiDaiNaiA Cai** = 'Be a BDNA to/of/about, etc. C'. So stress is an important feature of the speechstream in the neighborhood of letter-words.

As we've just seen, /VEDmabaicaiDAI/ will resolve as **Vedma Bai Cai Dai** without benefit of pauses. But if we put the stress on the middle term in such a set, we get something that appears at first sight--that is, before we rule on it--to be resolutionally ambiguous: /VEDmabaiCAIdai/. There's an acronym here, alright, but it might be either **BaiCaiDai** (/baiCAIdai/), or **CaiDai** (/CAIdai/) with the letter-word **Bai** coming before it. Clearly we must rule on this case. The best morphological ruling is to let the resolver take such a pauseless, penultimately stressed string of letter-words as the longest acronym it can be; and then use pauses to mark off other cases. Under this rule, the sense of 'Sell B to CD' is given by **Vedma Bai, Caidai**; and this will be uniquely resolved from either /VEDmabai.CAIdai/ or /VEDmaBAI.CAIdai/; but /VEDmabaiCAIdai/ uniquely resolves as **Vedma BaiCaiDai** and means 'Sell BCD'. Similarly, 'This is a ticket from B to C on DF for (price) G' is **Ti ketpi Bai Cai, DaiFai Gai,** which is uniquely given by /tiKETpibaiCAI.DAIfaigai/ (and some other variations). But again, only a single pause is obligatory.

2.31 Pause and Stress Around Dimensioned Numbers: There is a final problem involving acronyms, and that is how to protect the junctures around the acronym-bearing dimensioned numbers of Rule 8, Sec. 29. Our example was [100USD]. This is spelled out in text as [nemaUSaiDai], and partially spelled out as [100USaiDai]; but in any case it is produced in speech as /nema,uSAIdai/, the close-comma indicating that the /a,u/ is disyllabic. The problem is how to prevent these objects from being misheard as "indefinite descriptions", that is, with the same grammar as **Ne mrenu** = 'One man'. Like **mrenu, USaiDai** is a predicate. How is it that the resolver does not hear **Nema USaiDai**--which would mean 'Some one hundred instances of U.S. dollars

(i.e., things worth one U.S. dollar, e.g., the bills or coins themselves)'--in this production? Again, the default rule is to let the resolver hear the longest dimensioned number it can hear, and use pauses to mark off other cases. So the resolver is instructed not to resolve pauseless productions like /nema,uSAIdai/ as two words when it can resolve it as one. Indeed, the two-word phrase **Nema USaiDai** would be parsed as an indefinite description. So when indefinite descriptions involving acronymic predicates are intended, the speaker must pause between the quantifier and that predicate. Thus **Nema USaiDai** /NEma.uSAIdai/ gets the now-intended meaning of 'Some one-hundred U.S. dollars' exactly. Even consonant-initial acronyms require this protection. Thus, **Nema NaiZaiDai** = 'Some 100 New Zealand dollars' must also be pause-bearing (/NEma.naiZAIdai/) in order to distinguish it from the dimensioned number **nemaNAIZAIDAI** /nemanaiZAIdai/. This last expression is a single word, a quantifier, and might be the measure of some quantity, say '100 New Zealand dollars-worth of wool' = **NemaNaiZaiDai lunli.** No pause is needed in this indefinite description. The two stressed syllables in the pauseless production /nemanaiZAIdai-LUNli/ will effect the desired resolution.

Notice that /NEma.uSAIdai/ differs from /nema,uSAIdai/ in two respects: there is a pause and a stressed /NE/ in the first production but not in the second. The extra stress in the first production does help the human auditor resolve these two productions correctly; but it may not be relied on by the machine's resolver. For example, if the quantifier is monosyllabic, the natural tendency to stress one syllable relative to another vanishes. So /ne.uSAIdai/ requires the pause to distinguish it from /ne,uSAIdai/. Between this last pair of productions the phonological difference is now minimal; but it works.

One final note about stress in the neighborhood of dimensioned numbers. When such a number is of minimal length, consisting of only one letter-word, say, with a default **ne** assumed, and is being used in a numerical description--as in **Ti langa ta lio mei** = 'This is longer than that by (one) meter'--then the descriptor **lio** ('the number...') must be pronounced disyllabically and its second syllable must be stressed. This pattern will always give the desired resolution. Thus /liOMei/ will resolve as **lio mei,** while the ambiguous production */lioMEI/ is in danger of picking up any following letter-word as the final syllable of the number: /lioMEImei/ => **lio meimei** = 'the number (one) millimeter'. True; such a sequel is rare. But pronouncing **lio** disyllabically and stressing its final syllable offers complete protection against this accident; for then the one-syllable number is in effect the ultimate syllable of a quasi-compound, the phrase **lio mei,** which is in effect being stressed penultimately. This determines the right juncture of **mei** and prevents it from being heard as part of anything else.

2.32 Acronym Recovery Rules: The word-formation rules of Sec. 26 are sufficient to transform any (allowable) acronym into an acronymic word. But they are not sufficient to allow the recovery of every (allowable) acronym from its sound in speech. The consonant-words will decode uniquely, of course; but the single vowel-phonemes to which the vowel-words have been (nearly always) reduced will not. So to enable unique decoding of the single vowel phonemes in spoken acronymic words, certain already widely-used international conventions have been adopted. These are:-

a. **"Nationality":** The default assumption is that all the vowel-letterals in an acronym are Latin. If Greek vowels are used, they must be spelled out. If

all the letter-words aree single vowels, it is assumed that the entire acronym is Latin and upper case. Thus /aZAza/ is [AAA]; and the written word is **AzAzA.**

b. **Case:** It can be inferred from the word-formation rules that every single-letter vowel-word which is joined by a hyphen to a preceding upper-case consonant-word represents a lower-case Latin letteral. Thus /FAIze/ is [Fe] and the acronymic word is **Faize.** But let's go further. Let's also assume that every single-letter vowel-word that is not hyphenated to a preceding Latin consonant-word has the same case as that word, or as any following consonant-word if the vowel-word in question happens to be initial. Then /CAIo/ will decode as [CO] and /Unai/ as [UN]; and the two words will be spelled out as **CaiO** and **UNai,** respectively. (It is quite a different matter whether [UN] will regularly refer to the United Nations in Loglan, or to the (impossible) diatomic compound of Uranium and Nitrogen; we trust the former...although discussion of the latter is not of course impossible.) /FAIe/ of course will then be [FE]; /FAIze/ would still be [Fe]; and /FEIze/ (FAY-zeh) might well be taken to be [fE] if such a curious acronym is ever needed. (At the rate at which acronyms are proliferating in the modern world, it may soon be.) But /FEIe/ (FAY-eh) would always have to be written out as **feie,** for its acronym *[fe] would imitate the number-word **fe.**

The case convention will not allow the symbols for the two V-initial element words **Au** (Gold) and **Eu** (Europium) to be read aloud in the usual way. The problem is that both the letterals in just these two chemical symbols are vowels. So the above conventions would apply the default rule wrongly and decode the spoken words /AZu/ and /EZu/ as [AU] and [EU]. (One of these, [AU], as it happens, predicates another impossible diatomic compound, this one involving Argon. The other, [EU], is not a possible chemical acronym for [E] is not an element symbol.) So these two exceptional chemical acronyms will have to have their lower-case second vowels "spelled out" when spoken aloud, namely as /aZUsi/ and /eZUsi/, respectively. They can then be written [Au] and [Eu] as required. Elements whose symbols are single letterals are, of course, referenceable by the corresponding letter-words: thus Nitrogen by **Nai** and [N] and Oxygen by **Oma** and [O].

Interestingly enough, the symbols composed of single letter-words have the grammatical status of "arguments", that is, they function as designations; while acronymic words composed of two or more symbols--not necessarily all alphabetic--have, as we have noted, the grammatical status of predicates. Thus **Ta U235** /ta.utoTEfe/ means 'That's U-235'; while it takes "predification" (with Lexeme ME) to turn the argument **Uma** into a similar predicate: **Ta meUma** = 'That's U (in the sense of Uranium)'. The reason for this lack of parallelism between letter-words and acronyms is explained under Lexeme TAI in the Lexicon.

2.33 Resolving Structure Words: All structure words, whether simple little words or compounds, are resolved in a two-pass operation. On the first pass, the resolver **reduces** some part of some production that contains only little words to a string of simple little words, preserving whatever pauses and stresses it may find among them for use on the second pass. On the second pass, the resolver acts as a **compounder.** It places junctures (word-boundaries) around certain substrings of the string of little

words which the reducer has identified (in this way, it "compounds" them), thus isolating those which are left as simple little words. It then turns both kinds of "lexes" (words) over to the lexer for classification as "lexemes". (The terms 'lex' and 'lexeme' are defined in the next chapter; for the moment you can think of a lex as a word, and its lexeme as the part of speech to which it belongs.) Well-formed text has had all this morphological work done for it by the writer; so text is turned over to the lexer directly.

> [The "compounder" that will perform the second pass has not yet been written...although the lexer that will use its output has been. The latter is part of the preparser of the machine grammar which has, so far, been tested on textual input only. There is no doubt in my mind that a compounding algorithm to work with acoustic productions can be written, even if it might require a somewhat more elaborate array of pauses and stresses than are now thought to be sufficient. It is even possible that, to make the compounder work, some additional usage restraints on grammatical productions may also be found necessary; but these will almost certainly be minor adjustments in the usage rules which will have no large effects on the grammar.]

The **reduction pass** works as follows. For **unconditional resolution** the little-word resolver requires that it be given a segment of a production that is known by other resolutions of the resolver to contain only little words. Such a segment or segments could be (a) all of a V-final breathgroup which contains no CC's and hence no predicates; (b) those parts of a V-final breathgroup which does contain one or more CC's but in which the region or regions occupied by predicates have been marked (this will have been done for it by the predicate resolver in Sec. 2.60); or (c) a C-final breathgroup in which the regions occupied by predicates, if any, and by the resident name have all been marked, and in which there is one or more residual segments known to contain only little words. Thus in unconditional resolution, the little-word resolver is the last of the three resolvers to go to work on some breathgroups, namely on those parts of it which have been set aside for it by the other resolvers.

But there is a complication. The little word resolver is also required to perform **conditional resolutions.** The reader may recall that before the name-resolver of Sec. 2.17 could locate the left edge of the name that it knows is always resident in a C-final breathgroup, it had to have the "regular word resolver" (evidently a joint effort of the little word and the predicate resolvers) attempt to perform a conditional resolution of the prequel of any apparent name-marker that it found. This resolution was conditional because it could fail. Only when it found a name-marker whose prequel <u>did</u> resolve could the name-resolver mark off the region of the C-final breathgroup that was not occupied by the resident name. But not all prequels of apparent name-markers resolve. (The few that don't are, as we will see, easily identified.)

This seems circular but is not. The segment of a production which the name resolver calls a "prequel", and which is the string of phonemes that is given to the little word and predicate resolvers for conditional resolution, is definitely marked. In particular, it is the string that lies between the copy of a name-marker whose prequel it is and the left edge of that breathgroup. In fact the only difference between these prequel strings that the resolver is asked to resolve conditionally, and those more confidently marked portions of a breathgroup which are known to contain only regular words, is that the attempt to resolve the former may fail. The latter will resolve...if the utterance from which it is taken is well-formed. But the answer to the conditional

resolution question may occasionally be no. The resolver must be able to provide that sort of information as well as give the results of a successful resolution which it knows beforehand will succeed.

Let us commence with **unconditional resolution.** We will assume that the name- and predicate-resolvers have done their work, and that we now have a pauseless segment of a breathgroup that is known to consist entirely of little words. Such segments may be initial or non-initial in their breathgroups. We recall that strings of little words will consist only of V VV CV CVV elements; that V elements can only be initial in breathgroups or follow CVV elements; and that if a V element does follow a CVV element pauselessly the latter must be of Cvv form (as for example in the acronymic word **MaiA**); see Sec. 2.29. This detail is used in Step (3c) of the resolution procedure given below.

We proceed as follows:

1) We ask first if the segment is breathgroup-initial, that is, immediately preceded by a pause. If it is, continue. If it is not, go to Step (4).

2) If the first sound in the pause-preceded segment is V, the first word is either a V or a VV. To find out which, count the consecutive V's. If the number is odd, the first V is a word, and the remainder of the string of V's, if any, is composed of one or more VV words and are so resolved. If the number is even, the entire string of V's is composed of VV words and are so resolved. In this way, we arrive at the first C, if one exists in the segment, or at the end of the segment.

3) If the first sound (or the sound now being examined) in the segment is C, the next sound is a V; for, by construction, there are no CC's in this segment. So examine the next two sounds. Of the three possibilities, namely CVCV, CVVC, CVVV--CVCC is again impossible by construction--the first two resolve immediately:

3a) If CVCV, the first CV is a CV-word, and it is followed by the unresolved sequence CV... This returns us to (3).

3b) If CVVC, it can only resolve as CVV plus sequence C... (Recall that *CV + V is proscribed.) This too returns us to (3).

3c) If it is CVVV, we must listen to the first two V's and proceed as follows.

3c1) If they form a monosyllabic pair, giving CvvV, then Cvv is resolved as a word and we go with a V-initial sequence to (4), which handles V-initial segments that are not initial in their breathgroups.

3c2) If they are disyllabic, giving CV,VV, then this sequence can resolve in two ways. First, as CV + VV; second as a disyllabic CV,V + a V-initial sequence. To discover which, count the consecutive V's if any exist beyond CV,VV. If the number is zero or even, then the sequence resolves as a CV word followed by one or more VV words. If the number is odd, it resolves as a disyllabic CV,V word which is also followed by one or more VV words. (There can be no V-form words in the string of consecutive V's.) This brings us to the end of the segment or to the next C, which returns us to (3).

4) We know that the segment is not initial in its breathgroup. Therefore if it is V-initial, then it can only commence with a string of one or more VV elements; and these are so resolved. This takes us either to the end of the segment or to the first C; and the latter returns us to (3). If the non-initial segment is C-initial, we also return to (3).

In this way, all little words in well-formed utterances are unconditionally resolved.

We now consider the **conditional resolution** of a segment which is the prequel of a possible name-marker.

a) We ask first if the segment is C-final. If it is, it will not resolve into regular words and is so reported. If the segment is V-final, continue.

b) Is it breathgroup-initial? That is, is the segment preceded by a pause? Then any number of V's may lie between its C's, or before the first C, or after the last C. So resolve it unconditionally; for it will resolve into regular words.

c) Is it C-initial? If it is, every possible C-V pattern will resolve, so resolve it unconditionally. If it is V-initial, count the V's before the first C, if there is one, or in the segment if it contains no C. If there is an even number of V's, it will resolve unconditionally. It either is, or its head is, a string of VV-form elements. But if the number of V's is odd, it will not resolve and is so reported.

In sum, only if a segment is (i) C-final, or (ii) breathgroup-medial and commences with an odd number of V's will it fail to resolve. All other prequels of apparent name-markers will resolve unconditionally.

C. THE RESOLUTION OF PREDICATES

2.34 The Functions of Predicates: Predicates are the words with which claims are made and objects described in any language. Thus **Da mrenu** = 'X is a man' and **Ta blanu bakso** = 'That's a blue box' as well as **Le mrenu** = 'The man' and **Leva blanu bakso** = 'That blue box' essentially involve the predicates **blanu, bakso** and **mrenu** and 'is blue', 'is a box' and 'is a man' in each language. Predicates are also the most numerous words in any language, and in Loglan they belong to a single part of speech. But this is a grammatical matter which will be discussed more fully in Chapter 4. Morphologically, Loglan predicates are of two kinds, (a) the **little word predicates**--like the cardinals, ordinals, acronyms and identity words that have already been described as compound structure words--and (b) the **morphological predicates** that remain to be described. These are the words which are immediately detectable morphologically as predicates and are the subject of the present chapter. In this chapter the term 'predicate' will refer only to words which are immediately detectable as such unless it is specifically qualified to include little word predicates.

2.35 A Temporary Stress-Marking Convention: As we have already seen, stress is an important datum in resolving all varieties of predicates. So in this chapter, I will conventionally use an apostrophe ['] to mark the stressed vowel in most predicate words. Thus, **mre'nu** is merely a swifter way of writing **mrenu** /MREnu/. My use of this convention does not mean that stress is now to be marked with apostrophes in written Loglan. Quite the contrary. Stress in predicates is still regularly inferrable from the structure of the word, and so is omissible from the written form. It is to familiarize the reader with the inferrable pattern of stress among predicates that I shall be marking it here.

2.36 The Post-Emphatic and Intervocalic Pauses: Two occasions require that the speaker pause before a predicate. One is after emphasis; the second is when the predicate itself commences with a vowel. Neither is marked by a comma in text.

Occassionally an emphatic or stressed little word or syllable comes just before a predicate: /leTO.MREnu/ = 'The two men!'. If a pause were not used at this juncture, the production /leTOMREnu/ would resolve as **le to'mre nu.** At the present writing (1987) **to'mre** is not a word; but it is a legitimate word-form and would be heard as a new word by a novice or a computer.

Occasionally some non-initial predicate in the speechstream commences with a vowel. [None could in 1975; but some borrowed predicates now do.] Unless the preceding word is a name, which requires its own post-nominal pause, the preceding word will be regular and so end with a vowel. So in all such cases the required pause will be intervocalic. Intervocalic pauses are very brief, a mere glottal stop. Nevertheless, the pauses in **lo a'tmo** /lo.ATmo/ = 'atoms' and **le albumi** /le.alBUMi/ (leh . aal-BOOM-ee) = 'the albumin' are distinguishable from the same pairs of vowels pronounced pauselessly but disyllabically: /loATmo/ and /le,alBUmi/ (leigh-aal-BOOM-ee). These last productions contain audible glides between the adjacent vowels and would resolve not as phrases but as the single predicate words **loatmo** and **lealbumi.** Since these are also permissible borrowings they would be heard as such.

So it is important for the accurate resolution of predicates that both the post-emphatic and the intervocalic pauses be observed. Fortunately, they are both easy to observe, being quite common in natural speech.

2.37 Stress in Predicates: Stress is on the penultimate syllable in all predicate words including the numerical ones and acronyms. Some complex predicates like **saa'dja** have disyllabic terms; so the penultimate stress rule does not mean that stress will always fall on--or even in--the penultimate term. In **saa'dja** (saa-AAD-jaa) it does, or at least in it; in **matso'e** (maat-SOH-eh) = 'mother's-sister' it doesn't. Similarly, if **rie** in **casrie** is taken to be monosyllabic (as it may be; see Sec. 1.14), stress will fall on the first, or penultimate, term: (SHAAS-ryeh). If it is pronounced disyllabically--and currently it may be, at the speaker's option--then stress will fall on the first half of the last term: (shaas-REE-eh).

Neither hyphens nor buffering vowels are counted in fixing stress. Thus **mekykiu** is (MECK-uh-kyoo), and **matyma** is (MAAT-uh-maa).

Stress in other compounds and disyllabic little words has not yet been fully regularized...and indeed may never be, since as we have seen word-resolution does not depend on uniform stress production in these regions of the Loglan speechstream. (Will this remain a free area in Loglan morphology? Let's watch and see.) Thus, indicators

when disyllabic and initial in an utterance tend to be stressed on their second syllables (eigh-AA-moo-GOHT-soh) but need not be, and on their first syllables when alone or final: **Eo** = "Please|" is "EH-oh"; and **Mu gotso eo** = 'We go, please!' is "moo-GOHT-soh-EH-oh". But again, these patterns are not involved in resolution, and so are strictly optional.

2.38 The Forms of Predicates: Predicates are distinguishable from strings of little words by one major feature and two minor ones. The major feature is that, among the regular words of the language, that is, among non-names, all and only predicates contain adjacent consonants. One of the minor features is that, unlike little words, predicates may contain the sound /y/; where /y/ does occur in the predicates of unbuffered dialects (see Sec. 1.8) it is a hyphen. The other is that all predicates but not all structure words are stressed penultimately. Predicates share vowel-finality with the other regular words, the structure words.

Predicates may be of any length greater than three phonemes. Indeed any 4-letter or longer word is a "resolvable predicate"--although not always an allowable one--if it meets the following five requirements:

1. It must be uniquely resolvable as a single word.
2. It must be non-monosyllabic.
3. It must have at least one consonant-pair.
4. It must be penultimately stressed.
5. It must be vowel-final.

It will repay us to consider the first two requirements carefully.

First, nothing is a predicate that resolves as a phrase. Thus ***tokli'ni** is not a predicate because it is the phrase **to kli'ni**. Similarly, ***tosma'bru** is not a predicate because it breaks up into an operator, **to**, and something that <u>does</u> resolve as a predicate, namely **sma'bru**.

The second requirement is related to the first. Consider the monosyllabic psuedo-word ***spai**, once thought to be a permissible borrowing. But if ***spai** were a predicate, whenever it happened to be stressed relative to its sequel, it would join with the first V- or vv-final syllable of that sequel to create something that would always be another predicate. For example, suppose someone wanting to use ***spai** as a predicate said /SPAIleMREnu/ intending to say ***Spai le mrenu** = 'Spy on the man'. But what da's auditors will actually hear is **Spaile mrenu**; for **spai'le** meets all the requirements of a resolvable predicate. (Check them out.) Thus no monosyllabic predicate will always resolve uniquely. Whenever it was non-final and stressed, it would absorb the next syllable and create a word that would run beyond itself. Therefore there can be no monosyllabic resolvable predicates.

The three final requirements guarantee that we are talking about predicates and not about little words or names. That there should be at least one consonant-pair, and that stress be penultimate are the two discriminators of predicates in Loglan; and that they be vowel-final is a property of all non-names. But because of the intervocalic pause rule, that first CC need not be early in the word. Indeed, it may be among the last three letters of a predicate and preceded by an indefinite number of vowels. For example, **aoaoaoaoao'sti** (ow-ow-ow-ow-OWS-tee), though improbable, is a resolvable predicate. (Check it out.)

Suppose the speaker didn't say what da meant. Suppose one of those **ao**'s was

intended as an attitudinal, say the third one. Then all the **ao**'s to its left would also resolve as attitudinals, and the predicate would start immediately to its right, thus requiring that intervocalic pause: /aoaoao.aoAOsti/. Once da corrects da's error and the pause is spoken in this way, the new utterance resolves as **Ao ao ao aoaosti**; and **aoao'sti** is the resolvable predicate. Notice that no comma is used in text to announce this intervocalic pause.

The portion of a resolvable predicate that comes before its first CC (its prequel) must be composed entirely of vowels except that its first element may be a consonant. Thus **taoao'sti** is also a resolvable predicate.

There is one category of resolvable predicates which are not allowable. These are the borrowings that fail the "Slinkui Test"; of these ***sli'nkui** itself is an instance; see Sec. 2.58.

[The domain of possible predicate-forms has been hugely expanded over the limited 2 mod 3 sequences of 1975. Most of the new word-space is reserved for borrowings. Not only will this space be sparsely occupied, but as we will see later (Sec. 25), only some of that space is allowable.]

2.39 Three Kinds of Predicates: Predicates may be partitioned morphologically into three main types: **primitives**, which are of the two 5-letter forms [CV'C/CCV'] + CV; **complexes**, which have--unremarkably--a fairly complex pattern of forms which we will discuss presently; and **borrowings**, which are the residual of the other two. That is to say, anything is a borrowing that is an allowable predicate and not a primitive or a complex.

2.40 Primitives: These are the 5-letter words like **ma'tma** and **bru'di** ('mother' and 'brother'). There is no other form of primitive predicate. Some words that are morphologically primitives are derivationally borrowings. For example, **simba** (SEEM-baa) 'lion' is primitive in shape but was borrowed unchanged from Swahili, where it also means 'lion'. Similarly, the international word for 'meter' is **metro** in Loglan. This word too has been borrowed, but this time from the international scientific community where both the concept and some form of the word are all but universal.

Thus, derivationally considered there are **S-Prims** (taken from the international vocabulary of science), **N-Prims** (native or local words for things or concepts that originate in that place), **I-Prims** (international words, like **telfo** for 'telephone', that have spread internationally from a single origin), and finally, **C-Prims**, or those composite primitives which have been derived from as many of the eight target languages as possible and which are the "semantic primes" of the language.

2.41 Complexes: These are words like **furve'a** (foor-VEIGH-aa) = 'buy', derived from the phrase **fu ve'dma** = 'the one who is sold something'; **me'kykiu** (MEK-uh-kyoo) = 'ophthalmalogist', derived from **menki kicmu** = 'eye doctor', and **geoykre'ni** (goy-uh-KREH-nee) = 'gamma-ray'. Like compounds, complexes are composed of separately identifiable, meaning-bearing parts. These parts are called **affixes**. Unlike the parts of compound little words, however, all of which are intact little words, the affixes of which complexes are made are often, but not always, foreshortened versions of their source words. Thus **-kreni** is not foreshortened; but **mek-**, **-kiu** and **-vea** are. The

sources of the foreshortened words or affixes found in complexes are usually, but not always, primitive predicates. The sources of **mek-**, **-kiu** and **-vea** are primitives, namely **menki, kicmu** and **vedma;** but some are not. The sources of **geoy-** and **fur-** are the little words **geo** 'lower-case gamma' and **fu,** which is usually a sort of generalized passive. **Fu** has to be lengthened to become an affix; so the lengthener /r/ is used. Unlike /y/, which is a buffering hyphen--in **me'kykiu,** for example, /y/ buffers the two /k/s, which if allowed to come together would shorten the production to /MEKiu/, which would resolve as the little word string **me kiu--**/r/ is a kind of "morphological glue". It keeps things from falling apart. Without /r/ /fuVEa/ would resolve as **fu vea,** a meaningless little word phrase.

There is a **defining metaphor** standing behind each complex predicate. That metaphor or phrase constitutes the derivational source of the complex word. Thus, the metaphor behind **furvea** is the phrase **fu vedma,** which means 'the 3rd place of the predicate **vedma**'. **Vedma** in turn means '...sells...to...for price...'; so its 3rd place designates the buyer. From this we may infer that **furve'a** means '(to) buy'; and indeed it does. Behind **mekykiu,** as we have already learned, stands the phrase **menki kicmu** 'eye doctor' and behind **geoykreni** stands the standard international scientific metaphor 'gamma-ray'. Complexes thus come in a wide variety of shapes and sizes. All of them, however, resolve into a unique string of affixes with or without linking hyphens; see Sec. 2.46 for the contexts of hyphenation.

A final characteristic of every well-made complex of ***tosmabru**-shape is that it has passed the Tosmabru Test; see 2.56. (***Tosmabru** didn't; therefore we are obliged to star it.) There is more on complexes in Secs. 2.48-56.

2.42 Borrowings: This is the residual class composed of all the words that are resolvable as predicates--that is, that have the properties of predicates listed in Sec. 2.38--but which are not primitives and not complexes, and which have also passed the Slinkui (SLEENG-kwee) Test given in Sec. 2.58. It happens that all such words, or nearly all of them, have been borrowed from some natural language or group of languages. Thus **i'glu** is identifiable as a borrowing because it is a legitimate predicate that is neither a primitive nor a complex. Derivationally we know it comes from the Innuit (Eskimo) word of identical pronunciation. But even if we didn't know that, if we just looked at its sounds and sound-sequences alone, we would know that **i'glu** was a predicate of some kind. It has the required consonant-pair in /gl/; it is vowel-final; it is not a monosyllable; it is penultimately stressed (/IGlu/); and it cannot break up as a string of words or absorb parts of its environment to make a larger word (which is what words that fail the Slinkui Test unfortunately do). Furthermore, we know that **i'glu** is not a primitive because it is not like either **matma** or **brudi** in form. And finally, we know it is not a complex because it cannot be "parsed" into a string of affixes with or without hyphens; see Sec. 2.61. Some of these claims are certainly not obvious. But they are all true of **i'glu;** and your own developing morphological sense of the language will soon assure you that all these claims are true whenever they <u>are</u> true of some borrowing that you are considering putting into the language.

Derivationally, borrowings are as diverse as the languages and cultures from which they come. Thus **i'glu** comes directly from the languages of the Innuit, **alkurcu'fa** (aal-koor-SHOOF-aa) from the Arabic for artichoke; anthropology buffs will certainly be able to guess that **atla'tli** (aat-LAAT-lee) comes from the same Nahuatl word 'atlatl' (aht-ll-AHT-ll) for 'spear-thrower' that has been adopted by anthropologists for spear-throwers the world over; **pro'a** (PROH-aa) comes through English from the Malay

'perahu'; and **proto'ni** (proh-TOH-nee) imitates the Italian pronunciation of that word exactly but is really derived from all the European words for 'proton', including Italian 'protoni', inasmuch as they are all cognate borrowings of this Greek word for 'first'. Indeed the largest stock of borrowed words in Loglan comes from the international vocabulary of science. Thus words like **i'glu** and **atla'tli** are borrowed from natural language words in Loglan because cognate forms are widely used in an international scholarly discipline, namely anthropology; the Loglan word for 'fluorine' is **fluori'ni** (floo-aw-REE-nee) because of the nearly universal use of this word, or something very like it, in international chemistry; and **proto'ni** (proh-TOH-nee) is our word for 'proton' because, by what amounts to an international agreement, Greek 'proton' has been borrowed for this concept by all the cultures which do physics except those of China and Japan, in which a local metaphor of similar meaning is used.

There is more on borrowings in Secs. 2.57-58 and 2.62-63.

2.43 Consonant-Pairs: Any predicate word must have at least one CC. Long predicates often have several such pairs, and in some borrowings there are strings of 3 or even 4 adjacent consonants. Any adjacent pair of such consonants may be regarded as either an initial pair, like /br/ in **brudi**, or a medial pair, like /tm/ in **matma**. A different set of pairs is permissible in each position. The **permissible initials** are a proper subset of the **permissible medials**.

In an initial triple (e.g., /str/) both pairs (/st tr/) must be permissible initials. In a medial triple in a complex (e.g., /ntr/) the first pair (/nt/) need only be a permissible medial, but the second pair (/tr/) must be a permissible initial. In borrowings, both pairs in a medial triple (e.g., in /ntl/ both /nt/ and /tl/ need only be permissible medials.

2.44 Permissible Medials: Intelligibility studies of the consonant-joints have shown that the list of permissible medials that limited the predicate-forms of the 1975 language can be considerably lengthened. Any C_1C_2 is now permissible in mid-word unless:

	(a) $C_1 = C_2$	E.g.,	***kk**
	(b) C_1 is **h**		***ht**
	(c) C_1 is the unvoiced variant of C_2		***pb**
	(d) both are **c s j** or **z**,		***cs**
	(e) C_1 is **p t k** or **f** and C_2 is **j** or **z**		***pj**
or	(f) they are ***bj** or ***sb**.		

The need for the first two restrictions is obvious. The other exclusions are permutations which have been found to be frequently, or even always, unintelligible...even when medial. If any impermissible medial occurs at the C/C joint of a complex predicate, it must be hyphenated or avoided: **me'kykiu, cu'cysta, re'byjao**. Even hyphenated, however, the h/C permutations are proscribed.

2.45 Intelligibility at the C/CC-Joint: Complex predicates often have joints between a CVC-form affix and a following CCV-form affix, e.g., **m/br** in **ma'mbru** = 'mother's-brother'. The intelligibility of all possible joints of this kind has been studied. It was found that 19 unintelligible 3-letter permutations might occur at this joint were they

not proscribed. So these too must be hyphenated or avoided:

c/dz,vl	**d/cm,ct,ts**	**g/ts,zb**
j/dj,tc,ts,vr	**k,p/dz**	**m/zb**
n/dj,dz	**s,t/vl**	**v/ts**

Fortunately, all but **n/dj** are vanishingly rare in Loglan. **N/dj** occurs but can usually be avoided. Thus the hyphenated **sa'nydja** was avoided by using **saa'dja.** (***Sa'ndja** is heard as the apparent primitive **sa'nja** and may not be used.)

2.46 Hyphenation: Intraverbal hyphenation occurs in six contexts, all of them between the affixes of complex predicate words: (1) when the C/C joint in a complex is a proscribed medial (e.g., the **k/k** in ***mekkiu,** which must thus be **mekykiu;** see Sec. 2.44); (2) when the C/CC joint in a complex is proscribed (e.g., **n/dj** in ***sandja,** which must thus be **sa'nydja** (or **saa'dja;** see Sec. 2.45); (3) to attach a 4-letter affix to its sequel (e.g., **sanp-** to **-sensi,** which must thus be **sanpysensi** = 'sign-science'; see Sec. 2.51); (4) to keep an initial Cvv or CVV affix that is not representing a letter-word from "falling off" a word (e.g., **hai** from ***haihoi,** which must thus be **hairhoi** = 'happy-drinker'; see Sec. 2.55); (5) to attach a letter-word to the front of a complex which already has at least one CC (e.g., to attach **Xai** to **kreni** to make **Xaiykre'ni** = 'X-ray'); and (6) to spoil one of the "bridges" in a word that has failed the "Tosmabru Test" (e.g., the permissible initial **s/m** in ***tosmabru** itself, which must thus be **tosymabru,** or remade; see Sec. 2.56).

The primary hyphen is /y/; it is used in all of the above contexts except (4). /r/ is used in context (4) so that the complex that results from joining two or more CVV affixes will have at least one CC. In case an /r/ is already present at the V/C joint, its allomorph /n/ is used. Thus 'happily-angry' is **hai'nroi** /HAINroi/. 'Gamma-angry', if such a notion is ever required, would have to employ the full 5-letter affix of **groci,** yielding **geoygro'ci.**

2.47 Permissible Initials: The 1975 list of permissible initial CC's has actually been shortened. We will see later how this facilitates the recognition of borrowings. The list now includes only those pairs actually used initially in primitives. There are 36 of these:

bl	**br**	**ck**	**cl**	**cm**	**cn**	**cp**	**cr**	**ct**	**dj**	**dr**	**dz**
fl	**fr**										
gl	**gr**										
kl	**kr**			**jm**			**mr**				
pl	**pr**	**sk**	**sl**	**sm**	**sn**	**sp**	**sr**	**st**	**tc**	**tr**	**ts**
vl	**vr**									**zb**	**zv**

In 1975 permissibility far exceeded use. The only no-longer-permissible initial CC which was actually used in 1975 is **nr.** Its elimination was required to make an earlier version of hyphen **r** work, but with the addition of hyphen **y, nr** is possible again and could be added.

The restrictions on both initial and medial consonant-pairs apply only to predicates. In names, anything even vaguely imitative of the natural word is permissible: e.g.,

Xrucof, Strndl.

New permissible initials may be added from time to time; for example, **sf** is now being considered.

2.48 The Decipherability of Complexes: Deciphering a Loglan complex predicate like **mekykiu** is always possible because the affixes and hyphens of which complexes are made are always resolvable; that is, their joints can always be found out. Moreover, once an affix has been resolved in this way it can always be associated with exactly one predicate or little word, which means that the predicate can always be uniquely deciphered by the listener or reader who knows the meanings of all the affixes. Thus, while some predicates have been assigned more than one affix, no affix has been assigned to more than one predicate. Moreover letter-words, which imitate some affixes, never appear in the same contexts as the predicate-derived affixes. So if **mek** means **menki** 'eye' in one word, as it does in **mekykiu** 'eye-doctor', it will always mean **menki** in whatever other words it may appear. In this way, the defining metaphor underlying every complex can always be recovered.

The potentially ambiguous cases are those predicate affixes which happen to duplicate letter-words, mostly Greek ones, in current scientific usage. Thus **beo** is the CVV-form affix of the predicate **begco** = 'beg' or 'request'. But **beo** is also lower case beta, and is occasionally used in such technical terms as **beoytse'ro** (beigh-oh-uht-SEHR-oh) = 'beta-error'. By convention such letter-word affixes are always prefixes and always attached by hyphen /y/, not /r/, to words which are recognizable as predicates without them (as **tse'ro**, for example, is). In a similarly initial position a predicate affix would be attached by hyphen /r/. Thus, **beordo'u** (beigh-awr-DOH-oo) means 'beg-give' or '(to) grant a request'. CVV-form predicate affixes may also be final in complexes, where they bear no hyphen. Thus **gadbe'o** (gaad-BEIGH-oh) means '(to) pray', being derived from **gandi begco** = 'god-beg'. In sum, the letter-word or "beta" sense of **beo** is always expressed by **beoy-** (beigh-oh-uh), while the "beg" sense of **beo** is expressed in complexes either by **beor-** (beigh-awr) or by **-beo** alone. By marking them with a distinctive hyphen, the ambiguity that might otherwise arise from using letter-words in complexes is avoided.

The affixes currently assigned to predicates are shown in List 3. The affixes currently assigned to little words are given in Sec. 2.54. Any letter-words is a potential affix provided it is used prefixively and hyphenated with /y/.

2.49 Affix-Length and Frequency of Use: Two types of affixes are used in the construction of complexes: "long" ones and "short" ones. **Long affixes** are either 4 or 5 letters long; **short affixes** are all 3 letters long. By judicious selection of the lengths of the affixes the word-builder chooses, da can adjust the length of the resulting complex to be (roughly) inversely proportional to its expected frequency of use, thus satisfying the Zipf principle of bio-linguistical efficiency; see the reference to Zipf's work in **Loglan 1.**

2.50 Term-Reduction: When a complex like **ni'lboi** ('child') is composed entirely of short affixes (**nil** + **boi**) it is called **fully reduced.** The words in the defining metaphor of a complex (in this case **nilca botci**) are called its **terms.** To replace a term in a metaphor with a short affix is to **reduce** it; to replace it with a long affix is to leave it

unreduced. The set of short affixes has been designed to be an efficient "reduction machine". Currently, the efficiency of this machine is 95% in the exact sense that 95% of the terms of complex predicates in the 1975 dictionary could be reduced to short affixes if we wished them to be. The efficiency of the Loglan short-affix set--which is probably the largest in any language--will probably diminish as the semantic range and variety of complex predicates in the language increases. What this means is that complexes which are late additions to the language are less likely to be reducible than early ones...which is, of course, a fitting outcome since the early ones are those which are used more frequently.

2.51 Long Affixes: Every primitive has two long affixes. The one used in final positions is the 5-letter primitive itself, e.g., **-kre'ni** in **geoykre'ni.** The one used in non-final positions is the unique 4-letter affix that may always be obtained by removing the final vowel of a primitive. That is to say, primitives have been constructed in such a way that no two of them differ in only their final vowels unless they are members of the language/nationality/culture trios such as **spana/spani/spano** between which differentiation in a complex is seldom necessary. The place of that final vowel is then occupied by hyphen /y/. Take the metaphor **ni'lca bo'tci,** which means 'girl-boy', for the concept of "child" (not in the sense of "offspring", but the creatures who, along with puppies, play in village streets). Stripping the (non-distinctive) /a/ from **nilca** and replacing it with /y/ yields **nilcybo'tci.** So both or all affixes in a complex may be long. Or the first may be short and the second long: **nilbo'tci.** Or the other way round: **ni'lcyboi.** Or, as we have seen, both may be reduced. (**Ni'lboi** is in fact the dictionary entry for this sense of 'child'; it is derived from a Chinese-type metaphor, the logical sum of girls or boys.) Thus, every Loglan complex is in principle polymorphic, so its length may usually be adjusted to changing frequency of use.

2.52 Short Affixes & Their Derivations: The short affixes are all three letters long. There are four forms: CCV, CVC, Cvv with a monosyllabic vowel-pair, and CV,V with disyllabic vowels. For most purposes, but not for all, Cvv's and CV,V's may be considered variants of the same form. Examples of the four affix forms, showing all possible derivations from primitives, are as follows:

CCV: **cli** from **clika, fra** from **farfu;** rarely, **pre** from **papre, klu** from **kutla.**
CVC: **mat** from **matma, nil** from **nirli;** rarely, **rim** from **trime, tan** from **trana.**
Cvv: **boi** from **botci;** rarely, **lia** from **clina.**
CV,V: **cea** from **cenja;** rarely, **keo** from **ckemo.**

Affixes with the "best", i.e., the most memorable derivations--like **cli, mat** and **boi**--have been assigned to the most **productive** primitives, that is, the ones that appear in the greatest number of metaphors in the 1975 dictionary; and so their incidence in speech will hugely outnumber affixes derived by less obvious routes...like **klu, tan** and **lia.**

This fact may be usefully employed in constructing an optimal order for learning the short affixes in the affix-set. For example, if one has learned the most productive CCV-assignments first, and so has learned that **cli** belongs to the very productive **clika,**

then, when one comes upon the le uctive **clina**, one not only knows that its most obvious affix has been preempted an easily understand why. This makes psychological room, so to speak, for le **ia**.

This optimized order is the o hich the affixes are presented to the learner in **MacTeach 2.**

2.53 Affix-Assignment & Cover ot all primitives have short affixes. Currently the most productive 600-odd of t-odd C-prims have been assigned at least one short affix; and between 800 an short affixes have been assigned. Thus, some productive primitives have two or hort affixes for use in different contexts. For example, **madzo** has both **mad** and **enja** both **cen** and **cea**; and **godzi** has all three possible short affixes, namely **go** and **goi**.

Short affixes were assigned itives in such a way that their **coverage** of the metaphors in the 1975 lexicon w mized, that is to say, the number of terms in that pool of metaphors that was le by those affixes was maximized. Currently that coverage figure is 95%; see 2.50.

Short-affix space is extreme ed: 1,965 possibilities exist, of which the most useful 42% have already been assi Often what remains in the empty spaces of the assignment tables are forms like VE and ZVU which are unassignable. When this "unusable" space is removed--un that is, by the existing set of primitives, or anything like it--it turns out tha 75% of the usable space has already been used.

Moreover, there is often str petition between primitives for the remaining affix space. At the present, maage of affix assignment, such conflicts can only be resolved by remaking more pes. Affix assignment virtually stopped in 1982 when (1) coverage reached 95%,) the cost to old loglanists of relearning remade primitives reached approximately of the original 1975 list. This may have been premature. Perhaps 40 or 50 ad primitives should be remade, which would lift term-reduction efficiency by an % or 3%.

2.54 Preempted CVr Affixes: list of r-final CVC affixes was preempted for the 10 digits and 8 other opera und frequently in complexes. Currently these are: **fer for fur, jur, ner nir no ir, rar rer ror, ser sor, ter tor, ver vor** Apart from the digits they come, of co om the non-numerical quantifiers **ra re ro**, the negative **no**, the decimal point **pi** he three simple conversion operators, **nu fu ju.**

2.55 Making Complexes: To ma omplex predicate, one should first gauge the expected frequency of use of th pt to be rendered. Then a metaphor must be contrived that allows the degree c-reduction necessary to produce a word of the length appropriate to that freq rank. [The Institute will eventually publish length/frequency tables for the gu of word-makers.] Suppose a complex is to be fully-reduced. The rules for com short affixes are as follows:-

1. CCV's may go anywhere px; they never require a following hyphen.
2. CVCs may be only non-f the consonant-joint with the next affix is impermissible, they must henated with /y/. E.g., **mekykiu** and **keryrou.**
3. Cvv's and CV,V's may go re, but if initial in anything but a 2-term word with final CCV, e.g., **boi** y must be hyphenated with /r n/.

As an illustration of Rule 3, if it is desired to combine **boi** + **cli** + **mre** ('boy-like-man'), an /r n/-hyphen must be introduced between the first 2 terms: **boircli'mre** (boy-rr-SHLEEM-reh). Otherwise, **boi** would "fall off". Without the hyphen, /boiCLImre/ would resolve as the phrase **boi climre** = 'beta is a "similar-man" (whatever that means)'; and it doesn't matter that the sequence **cli'mre** is currently unassigned. To the computer's ear--and that of the novice--this is an unknown complex with a decipherable metaphor, "similar man".

Similarly, in a 2-term word composed entirely of Cvv's and/or CV,V's, e.g., **hai** + **hoi**, the joint between them must also be hyphenated with /r n/: **hairhoi**. Otherwise the production /HAIhoi/ is just a pair of little words. There is more on hyphenation in Sec. 2.46.

Long affixes may always be substituted for short ones. If the final term is to be left unreduced--for example, if greater length is required, or if there is no usable short affix for that primitive in that position--then its 5-letter affix may be used. If a non-final term is to be left unreduced, then its 4-letter affix plus hyphen /y/ may be used.

Each trial word must then be tested for resolvability.

2.56 The "Tosmabru Test": Only one type of complex word has known resolvability problems. These are the "Tosmabru-type" words in which the last term is a long affix of **ma'bru**-form in which the medial consonant-pair is (like **br**) a permissible initial. All non-final terms (there may be several) must be CVC's (like **tos**) for resolution problems to arise. Tosmabru-type words can break up into phrases...as indeed /toSMAbru/ does: **to smabru.**

To test your trial word for "Tosmabru failure", examine all the C/C joints between the CVC affixes, and between the last CVC and the **mabru**-term. If all those C/C joints are **bridged** by permissible initials, then the word will break up. But if any C/C joint is **unbridged**, i.e., is impermissible as an initial CC, your word will not break up. It has passed the Tosmabru Test. Interestingly enough, only one unbridged joint per word is sufficient to ensure resolvability.

Suppose your trial word is **?gusnilbo'tci.** Its intended terms are **?gus/nil/botci.** S/N is bridged; L/B is not bridged. One unbridged joint is enough. Therefore it passes. The production /gusnilBOTci/ cannot resolve as the phrase **gu *snilbotci** because ***snilbotci** fails the "Slinkui Test", which is a test for borrowings; see Sec. 2.58. (Before we've tested it **?snilbotci** is an apparent borrowing.) Suppose our trial complex is **?gusnitca'pri.** Its intended terms are **gus/nit/capri.** Both S/N and T/C are now bridged. Therefore ***Gusnitcapri** fails. It is in fact the phrase **gu snitcapri. Sni/tca/pri** is clearly quite a different complex.

Tosmabru failures may be repaired in several ways. One is to try different affixes, or even different metaphors, until a joint appears that is not bridged. Another is to hyphenate. Hyphenating the first C/C joint in ***gusnitcapri** (but not the second) would repair it. Thus **gusynitcapri** (goos-uh-neet-SHAAP-ree) resolves. There is also a simpler solution in this case. **Gusto**, the obvious source of **gus**, has another short affix, namely **gut.** Using **gut** instead of **gus** in ***gusnitcapri** demolishes the first bridge, and so **gutnitcapri** resolves.

2.57 Allowable Borrowings: There are four tests which must be passed by a usable borrowing. Any word which is (1) a resolvable predicate (Sec. 2.38), (2) neither a

primitive nor a complex, (3) free of "imitation hyphens" (it obviously can have no real ones if it is not a complex), and (4) neither a **Slinkui-type** word nor a **left-extension** of such a word, e.g., ***aslinkui**, is an allowable borrowing.

***Slinkui** is a disallowed word, typical of its class, that on the surface seems perfectly permissible as a borrowing. It is certainly a resolvable predicate. It is certainly not either primitive or complex. But then one notices that if ***slinkui** were allowed in the language, a certain perfectly regular complex, namely **paslinkui** (which means 'ancestor', from 'past-linear-kin'), would not uniquely resolve. For if ***slinkui** were allowable, **paslinkui** would break up as the phrase **pa *slinkui**.

On the other hand, if we dissallow ***slinkui**, and all words of its type, then **paslinkui**, and any word of its type, will not break up. The resolver would first obtain ***slinkui**; note that it was disallowed; look at the prequel; find **pa**; and then it would say, Right! **Paslinkui** must be it, then. That's the only thing /pasLINkui/ can be, since ***slinkui** is not allowed. (This, incidentally, is how **Aslinkui-type** words would also get resolved...if we allowed them. But these left-extensions of the basic Slinkui-form are disallowed for a slightly different reason which I will take up presently.)

The principle underlying the exclusion of Slinkui-type words from the domain of borrowings (some apparent Slinkui-words, by the way, e.g., **smabru**, are quite acceptable as complexes) is that complex predicates are so much more important in Loglan semantics than borrowings are--it is only the imitativeness of borrowings, after all, and not their number, that we intend to enhance by giving them a large word-space--that whenever there is a competition for a given portion of the word-space between these two varieties of predicate, it is the complexes that must always win out.

Thus, Slinkui-type borrowings (but not complexes) are not only excluded from the class of allowable borrowings, they are excluded from the language. That is why the "perfect complement set" I once thought we could achieve for borrowings is not perfect. It has a big empty swath cut right down the middle of it, namely the space occupied by all those resolvable predicates that are neither primitives nor complexes yet fail to pass the "Slinkui Test"; see below.

A second, similarly large emptiness in the domain of resolvable predicates is caused by our disallowing Aslinkui-type words as well. While these words are definitely resolvable once Slinkui's have been disallowed (but not before), the great variety of those possible leftward extensions of any Slinkui-form makes them difficult to see or hear. Indeed, they impose a "double double-take" on the resolver, a twisted loop in the resolution algorithm. So for these reasons--or this reason, if it is the same one--Aslinkui's too have been disallowed.

Let us finally consider the reasons behind the third test for allowable borrowings. We said it must contain no **imitation hyphens**, that is, no **y**- or **r**-involving sequences that imitate the hyphens of a complex. This exclusion is not always strictly necessary for resolution. It is not necessary, for example, when the word containing the apparent hyphen will not resolve as a complex. But in anticipation of the day when borrowings may also be involved in complexes--when some Innuit-speaking loglanist invents the word **i'glymao** for 'igloo-maker', for example--in which case hyphenated segments in a complex may be exceedingly irregular, let us now exclude all those imitation hyphens from "simple" borrowings that might confuse such constructions later. What this means is that no /y/, no /r/ in the context /CVVrC/, and no /n/ in /CVVnr/ is permitted in a borrowing.

2.58 The "Slinkui Test": Any CC-initial trial borrowing must be given this test; no

other need be. It is simply performed. Put a test prefix, say **pa**, on your trial word. Let's say it's actually **?slinkui**. Then try to resolve **?paslinkui** as a complex; see Sec. 29. If the effort succeeds, as it does with **pas/lin/kui**, then the trial-word fails. If the term-resolving effort fails, as it would, for example, with **pas/lin/kuti**, then no coalescence with a preceding CV operator is possible; so the trial word passes.

Suppose your trial word is **?spe'a**. **?Spea** has a form that was once accepted as an allowable borrowing. But **?spe'a** obviously fails the Slinkui Test. **Paspe'a** immediately resolves as complex and so ***spea** fails as a borrowing. [Thus a whole set of words, once accepted as legitimate borrowings--albeit in a narrower, modularly defined lineage--now vanishes from the domain of allowable borrowings. And so from the language.] But trial words that fail the Slinkui Test can usually be easily repaired. Any disturbance in the smooth structure of the imitation complex that takes over after the first letter--e.g., the **linkui**-part of ***slinkui** or the **pea**-part of ***spea**--will also suffice to make it pass the test. For example, ***sli'nkui** could be repaired as **slie'nkui**, **slinkui'a**, **slini'kui**, or, as noted, **slinku'ti**. Indeed, dozens of variations on each trial word are usually possible; one of them might imitate its natural model almost as well as the word that failed.

Some Slinkui-form sequences discovered by the resolver in the course of word-resolution are actually complexes. For example, if one presented **?sma'bru** to the Slinkui Test, one would find out it was a Slinkui. No question of "failing" as a borrowing is involved here...merely the identification by the resolver of certain legitimate sequences as having Slinkui-form. There is more on this diagnostic use of the Slinkui Test in the next two sections.

2.59 The Resolution & Partial Classification of Predicates: Under the new morphology, the resolver proceeds very much as before, except that it must now deal with the possibility of vowel-initial predicates and it must perform the Slinkui Test on any CC-initial predicate-form sequence it encounters. Let us see why.

Suppose some such sequence "fails the test", that is, turns out to be a Slinkui. Since Slinkui-type borrowings are not allowed, and therefore do not exist in any well-formed string the resolver will encounter, the discovery that it has found something that looks like one--that this particular CC-initial sequence is indeed of Slinkui-form--can only mean (1) that it's a complex (e.g., **sma'bru**), or (2) that the sequence is part of some larger predicate. Slinkui-form sequences can only exist as words if they are complexes; so if some sequences exist that are not complexes, they must be parts of words. The kind of word of which they are parts must obviously be predicates. Moreover, the part they are is a right-hand part since the right juncture of the Slinkui-sequence will already have been fixed; see next section. Therefore there must be a left-hand part out there somewhere.

What kind of a left-hand part could it be? Well, if weren't for the Aslinkui exclusion, it could be anything that could precede an impermissible initial consonant pair that turned up as the first CC of a predicate. For in truth, the **sl** of a ***slinkui** that is not complex cannot begin a word. Since Aslinkui-type borrowings have been excluded from the language as well (see above), there is only one allowable form for the left-hand part of that larger word, and that is the CV-form that will by definition turn any Slinkui into a complex. There will always be one, of course, or the non-complex Slinkui would not be there. So the resolver looks for the CV ahead of each such Slinkui; always finds one (in any well-formed utterance); and puts left-juncture to the left of that CV and hence the word.

The word that results from ver will always be a well-formed complex; the Slinkui Test will have seen to t sense, the resolver short-circuits the normal recognition procedure in thes cases. For normally the recognition of a complex is subsequent to its r as a predicate.

2.60 The Predicate Resolution: The new algorithm for resolving predicate words is in some ways simpler 1975 one. The resolution of a predicate still commences, of course, with th y of a CC in the left-to-right scan of some breathgroup. That first CC ay be others, of course--may be either a permissible initial (**sp**) or missible one (**rk**). If the FIRST CC is IMPERMISSIBLE, the algorithm that fact as a prevailing condition and then performs the following steps:-

1. Is there a stressed V' mediately ahead of the first CC? If so (e.g., **i'sp, ia'rk**), mark it a STRESS FIRST as a prevailing condition. If not, find the first ins V' or vv' that follows the first CC (e.g. **spe', spudei', rkia', rkoti'**) it.

2. Find the first instanc or vv after the marked V' or vv' (**i'spai, ia'rko, spe'a, spudei'tai, rkia'**). Put right juncture to the right of that V or vv (**i'spai|, ia'rko|, spudei'tai|, rkia'mpa|**, etc.).

3. If the condition STRE and/or FIRST CC IMPERMISSIBLE obtains, find the prequel by locat irst instance of a C or /./ to the left of the first CC (**.aoi'spai|, tia'rko'mpa|**). Put left-juncture at that /./ or before that C (**|aoi'spai|, |t morkia'mpa|**) and exit.

4. Perform the Slinkui T sequence that remains. (Stress will follow its first CC, and it will issible initial.) If it is a Slinkui (e.g., **spe'a|, spa'rgu|, sma'bru|**), s he term resolver to discover whether it resolves as a complex. If it do resolve (e.g., **sma'/bru|**), put left-juncture immediately to its lef **ru|**). If it doesn't resolve as a complex, find the CV that will alwa ts left and put left-juncture to the left of that CV (**|tospe'a|, |mispa**). In either case, tag it as a recognized complex and exit. If it is no i (e.g., **spudei'tai|, spu'ta|**), put left juncture immediately ahead of CC (**|spudei'tai|, |spu'ta|**) and exit.

Clearly term-resolution is invo least some cases of word-resolution. We must now consider how the resolver w " the terms of a complex when the predicate is one.

2.61 Term-Resolution: Once te has resolved as a word, the recognition of its primitiveness, if it is primitiv ial: it is either of **matma-** or **brudi**-form, or it is not. If the resolved word imitive, or if a Slinkui-form sequence has been identified by the resolver, an is made to parse the word or sequence as a complex, that is to say, to reso erms if it has terms. If it has no hyphenation sequences and is 0 mod 3, then, i mplex, its terms will just "fall out" in 3's, and each triple will be a legitimate **ma/bru, nil/boi, tar/ses/mao, roj/mad/ses/mao,**

and so on. If they don't or aren't (e.g., **tai/rko, mor/kia/mpa**), it is not complex. If it has no hyphens and is not 0 mod 3, then it will have a 5-letter final affix if it is complex; and the rest will fall out in 3's. If they don't and/or there is no long final affix, it is not complex. If it has hyphens or hyphen-like sequences, e.g., /y/, /CVVrC/ or /CVVnr/, it is complex. (Borrowings, remember, may contain no imitation hyphens.) So the algorithm then examines the regions between the hyphens, and between any hyphen and the ends of the word, and the lengths of these regions, together with their positions in the word, will always uniquely determine what lengths of affixes will be found in them if the complex is well-formed. (It will also identify whatever irregularly formed sequences it finds among the affixes as derived from borrowings, when this development is upon us.) In this way, every predicate word that is a well-formed complex will uniquely resolve into its terms; and every word that resolves into terms will be a well-formed complex.

2.62 The Recognition of Borrowings: The recognition of a borrowing thus always follows on the failure to hear or see it as a complex once it has resolved as a word. Thus, one knows one has a borrowing only by failing to resolve it as a complex. Thus **tia'kro** is complex because its two affixes, **tia** and **kro**, are well-formed. **Tia'rko, morkia'mpa** and **fainstoi'a** are not complex because their partition as complexes fails. **Rko** and **mpa** are not well-formed CCV-affixes; and **fainstoi'a**, while 0 mod 3 and hyphenless, does not fall out into legitimate 3's: **fai/nst/oia.** And so all these words are borrowings.

The construction of a borrowing thus often proceeds by putting impermissible initials (**rk mp ns**) in judicious places, or by making sure that its affix-length pieces do not have affix-form shapes (**nst oia**). Thus, we may say that whatever is an allowable predicate word and not a primitive or a complex is a borrowing.

Observe, by the way, the operation of the pair-from-the-left rule of Sec. 1.15 in determining the character of the stressed syllable of **fainstoi'a:** /fainSTOIa/ (fighn-STOY-aa), which is quite a different sound from (fighn-STOH-yaa).

2.63 Making Borrowed Predicates: The procedure for making a predicate based on some natural language model involves four distinct steps:

1) Build a preliminary trial word that imitates the model word as well as possible...either its sounds or its letters, or rarely, both. When the model predicate is a scientific word, the word-maker may be guided by the transcription system given in Sec. 2.14 for pronouncing Linnaean names. Thus the predicate 'cephalopod'--which is either derived from or the source of the Linnaean name **Cephalopoda**--might well be initially transcribed as **?cefalopo'd**; for this is our recommended pronounciation, and hence a rewriting in Loglan phonemes, of the portion it shares with the related Linnaean name. Even if there is no related Linnaean term, transcribing it as if it were Linnaean is a good procedure to follow when borrowing scientific words.

2) Make sure the trial word is a resolvable Loglan predicate, repairing it as necessary to give it the necessary properties; see Sec. 2.38. **?cefalopo'd** (sheh-faa-loh-POHD) lacks two of these properties. It is not V-final and it has no CC. Let us repair the first defect by giving it the final **-a** of the corresponding Linnaean name, and the second by inserting **r** before the second C. /rC/ is always an impermissible initial,

and having one sufficiently early in the word is crucial for preventing the first C and its following one or more V's from "falling off the word". **?cerfalopo'da** (shehr-faa-loh-POHD-aa) is the word that results from these moves.

3) Make sure that the developing word is not primitive or complex, and that it doesn't resolve as a phrase. This means checking its form against the two forms reserved for primitives--which require a more cautious procedure of derivation than you may wish to get involved with here--and attempting to resolve it first as a phrase, and when that fails, as a string of affixes. Both attempts fail with **?cerfalopo'da.** Nothing falls off; and term resolution starts out with **cer/fal/..** but then encounters a sequence which is impossible in a complex, namely **...opoda.** So the word passes the third test.

4) Check it for false hyphens. **?cerfalopo'da** has no /y/s, no /CVVnr/s, and the /cerf/ sequence just misses being an instance of /CVVrC/; so the developing word passes this test as well.

5) Test it for Slinkui Failure and the Aslinkui Condition. The first test is relevant only if the trial word commences with CC, as this one doesn't. The second condition can exist only in a trial word whose first CC is non-initial, is a permissible initial, comes before the stress, and has a prequel that resolves as words. Only in such words will the prequel be loosely-attached. The first CC, namely /rf/, is not a permissible initial; so the trial word avoids the Aslinkui Condition as well.

Cerfalopo'da passes all tests and is thus an allowable borrowing. Eventually, for at least the borrowed portion of the Loglan scientific vocabulary, this word-borrowing procedure can easily be made algorithmic. The borrowing algorithm is under development now. But much of the Loglan scientific vocabulary will not be borrowed. Large portions of it may be captured Chinese-style by local metaphor. For example, rather than borrow the sounds of 'xenophobe' one might prefer to "borrow" its idea, the idea of the "stranger-fearing" person. That would produce a scientific complex that would be decipherable by <u>any</u> Loglan reader, not just those learned in the vocabulary of Western science. If this were our policy, **gu'rfia** might well become the word for 'xenophobe/-ic' in Loglan. It is made from **gutra firpa** = 'stranger-fear' and is thus a literal translation from the Greek. Whence **lopo gu'rfia** would be 'xenophobia', and so on. Still, the borrowing **zernofo'bi**--made by much the same procedure as **cerfalopo'da** was made above--is also possible, and has the advantage of being immediately recognizable to those who know the psychiatric vocabulary. But even in English--one might say especially in English--such Graeco-Latin words exclude from understanding precisely those whose educations have not permitted them to know.

To borrow or not to borrow. Lehnwort vs. lehn übersetzung. There are many good arguments on both sides of this question. As yet, The Institute has taken no position. We prefer to wait until the loglaphone community has extended itself internationally--and also intellectually, one would hope--beyond the narrow confines of anglophone computer science and those related Western disciplines of logic, linguistics and anthropology where interest in Loglan began some decades ago.