

Proposed Reference Grammar of the Loglan Language

M. Randall Holmes

May 11, 2018, 2:38 pm Boise time

Contents

1	Introduction	7
1.0.1	Version notes	8
1.1	Notable changes from 1989 Loglan, in brief	9
2	Phonetics	11
2.1	Sounds	12
2.1.1	Vowels	12
2.1.2	Consonants	13
2.1.3	Vowels pretending to be consonants	13
2.1.4	Consonants pretending to be vowels	13
2.2	Diphthongs and vowel grouping	14
2.2.1	Mandatory diphthongs	14
2.2.2	Optional diphthongs	14
2.2.3	Expanded notes on diphthong pronunciation	15
2.2.4	Grouping long streams of vowels in predicate words and names	16
2.2.5	Grouping long streams of vowels in structure words	17
2.2.6	Vowel pairs with optional grouping revisited	17
2.2.7	Doubled vowels and stress	17
2.3	The Loglan syllable	18
2.3.1	Discussion of stress and notation for syllable breaks and degrees of stress	18
2.3.2	Every syllable has a vocalic unit	19
2.3.3	Like Gaul, every syllable has (up to) three parts	19
2.3.4	Words must resolve into syllables	20
2.3.5	Notation for consonant and vowel patterns	21
2.4	Word Forms	21
2.4.1	Word forms enumerated	21

2.4.2	Pauses and word boundaries	21
2.4.3	Structure words (cmapua)	22
2.4.4	Name words and the false name marker problem	25
2.4.5	Borrowed predicates	27
2.4.6	Complex predicates	29
2.4.7	Appendix: Alien Text and Quotations	33
2.5	There is a rationale, or "The original sin of Loglan?"	34
3	Lexicography	39
3.1	Name Words (and Alien Text)	40
3.2	Predicate Words	41
3.3	Structure Word Classes	43
3.3.1	Tight logical connectives: CA roots	43
3.3.2	Letters, acronyms, and pronouns	44
3.3.3	Remarks on acronyms	46
3.3.4	Numerals and quantifiers	47
3.3.5	Tense/location/relation operators	51
3.3.6	Connectives	59
3.3.7	Articles	64
3.3.8	Constructions involving alien text and related articles (see the appendix to the Phonetics Proposal for some modifications)	65
3.3.9	Assorted grammatical particles, somewhat classified	67
3.3.10	Words which form free modifiers	70
3.3.11	Negation	74
3.4	Essays on word-making, and on what a word is exactly	74
3.4.1	Borrowing predicates	74
3.4.2	Making complex predicates	75
3.4.3	Name words	77
3.4.4	Essay: what is a word?	78
4	Grammar	81
4.1	Introduction	81
4.2	Sentences, Free Modifiers, and Utterances	81
4.2.1	The basic SVO statement	82
4.2.2	Subject-final statements and observatives	83
4.2.3	Sentences	84
4.2.4	Note on right closers	87

4.2.5	Free Modifiers	87
4.2.6	Utterances	90
4.3	Predicates	92
4.3.1	Verb Phrases	92
4.3.2	Predicates	95
4.4	Arguments, Modifiers, and Term Lists	98
4.4.1	Predicate Modifiers	98
4.4.2	Serial names	98
4.4.3	Arguments (and subordinate clauses)	99
4.4.4	Term lists and termsets	103
4.4.5	Linked term sets with je/jue	104
4.4.6	The Alternative Parser, summarized	105

Chapter 1

Introduction

This document contains a proposal for a complete description of the grammar of the Loglan language as promulgated by the Loglan Institute, presented for consideration by the Academy and the membership.

There are some things it does not include for which one should consult other resources. The dictionaries in the HTML version are indispensable: I suggest using my latest versions.

It contains no discussion of the process by which primitive CVCCV and CCVCV predicate words were originally generated. We do not expect this process ever to be used again by the Institute: we simply regard these predicate words as being among the primitive word roots of Loglan.

We also make no mention (or little mention) of the Sapir-Whorf hypothesis.

Loglan 1 (1989) and Notebook 3 remain valuable resources for understanding how the language got to where it is.

Exact details of how the grammar works can be discovered by testing with the latest parser versions. On the other hand, bugs in the PEG grammar may on occasion be detected by comparing what the parsers do with the intentions laid out here. The PEG grammars, `draft-grammar-with-comments.peg` and `draft-grammar-with-comments-alternative.peg`, are in principle a complete description of my intentions in a different mode, and are extensively commented in English.

Extended discussion of motivation of changes to the grammar proposed here are not usually present here. They can be found in the Report document.

Loglan grammatical terminology is an important issue. The reader should notice that I am using somewhat different terminology in English than is

traditional in our sources. Eventually I would like to have an adequate suite of terms in Loglan for discussion of Loglan phonetics, lexicography, and grammar: developments along these lines will be recorded here.

This remains a work in progress.

1.0.1 Version notes

5/1/2018: This is the first posted version. It still contains some notes about points in the grammar that may need fixes (I will resolve each of these either by proposing something or doing nothing and removing the comment). It needs a lot more examples, throughout. Readers who have the parsers are welcome to submit examples of the phenomena described, or counterexamples which suggest that things may not be working as described!

5/1/2018: Noting a systematic edit: I went through and made sure that references to predicate *words* were made in a way distinguishable from references to *predicates* (grammatical structures). I also added a final section on underlying principles to the phonetics chapter: it is my view that the phonetics unfold with some inevitability from a fairly small number of decisions made by the Founders, but this may not be obvious from the extended description.

5/1/2018: Some error corrections: fixed an incorrect statement about negative arguments appearing in several places and fixed a defect in the section numbering.

5/10/2018: using “adverbial modification” instead of “verb modification” for what JCB calls “metaphor”. This can readily be reversed.

5/11/2018: added an appendix describing the “alternative parser” innovations in one place. Executed various grammar corrections suggested in footnotes: the footnotes are still in place with the remark that they are implemented.

1.1 Notable changes from 1989 Loglan, in brief

We give a list of obvious changes (the sorts of things one runs afoul of when trying to parse older Loglan texts). Details can be found in appropriate parts of the grammar. This list is not exhaustive, and there are a fair number of subtler changes.

There is no pause/gu equivalence: There is no use of pauses as right closers.

Names: Loglan names are always spelled phonetically. Continuants must be doubled, which causes changes of spelling in some names in old texts. A name cannot end in three consonants: this has so far always been correctable by doubling a continuant. The old Linnaean article **lao** is used for names with foreign spellings.

Strong quotation: The new strong quotation is quite different. It is modelled on the 1990's treatment of Linnaeans.

Acronyms, pronouns and letterals: Acronyms are names, not predicates, and acronyms used as dimensions must start with the marker **mue**. There are no multiletter pronouns and there is no need to pause between letterals occurring as successive arguments.

Observatives: Something which would have been a tensed imperative in 1989 Loglan is now an observative (a sentence with indefinite subject understood, like **Na crina**, "It is raining" (literally, some unspecified person or thing is being rained on)).

Subjects: Only one argument without a case tag is permitted before the predicate in an ordinary SVO sentence (second and subsequent such arguments must be separated from the first by a new particle **gio**) and a gasent must have only one untagged argument after the second **ga** or all arguments (in fact all terms) after the second **ga**.

No unmarked vocatives: Vocatives must be marked with **hoi** or one of the words of social lubrication.

Alternative closers for abstract descriptions, subordinate clauses: Special suffixed forms are designed to allow fewer closers to be used. In

general, right closers often work a little differently than in 1989 Loglan, though simple examples should usually work in the same way (but not necessarily with the same parse).

One must pause after IPA, APA: One must pause explicitly (comma written) after these connectives or take the new approach of closing them with **-fi**.

Chapter 2

Phonetics

In this chapter I present the phonetics of Loglan and the strictly phonetic aspects of the rules of word formation for words of the various classes present in the language. This chapter is perhaps the most conservative in this document. It is more precise and thorough than the treatment in Notebook 3 and Loglan 1, but it makes very few changes. It does make the substantial addition of supporting explicit notation for syllable breaks and stresses, making possible a phonetic transcript style which the parser can handle. Thus we make no use of JCB's phonetic notations, having our own.

Some of it is quite baroque and could be improved in various ways (some of it perhaps *should* be improved), but my present mission is to get an official view of what the current state of affairs is. Essentially everything in this chapter is about phonetics and word forms: there are allusions to other perhaps controversial features of my provisional grammar (such as converting acronyms to names) but these have no bearing on the phonetics: they will be covered in detail in later chapters.

There is concrete evidence for my claim that this is conservative. I parsed *every word in the dictionary*. Only a handful needed to be changed, and most of those were wrong in 1989 Loglan terms.¹

¹Parsing every word in the dictionary has now been carried out with the PEG implementation of the new proposal. There were no new changes needed that had not already been seen to be needed in earlier parser versions.

2.1 Sounds

2.1.1 Vowels

The Loglan vowels are the *regular vowels* **a,e,i,o,u** and the *irregular vowel* **y**. The pronunciations of the regular vowels are typical Continental European (not English!) pronunciations.

Vowels appearing singly (not adjacent to another vowel) are pronounced as follows:

- a** is pronounced as in *father*
- e** is pronounced as in *bet*
- i** is pronounced as in *mach*ī*ne*
- o** is pronounced as in *lost*
- u** is pronounced as **oo** in *poor*

All of these are pure sounds. They can generally pronounced as in most languages spoken in continental Europe (English is severely aberrant in its spelling).

The sound of **y** is officially the schwa (the unstressed vowel in English *data* or the stressed vowel in English *hunt*), but we think there is something to be said for it being another sound easily distinguished from the regular vowels, such as the *oo* in English *look*.² We have also considered ö and the Cyrillic letter that looks like бI³ as implementations of Loglan **y**. English and Russian speakers must be cautioned against pronouncing unstressed vowels such as the **a** in **matma**, *mother*, as a schwa: these vowels should have the identifiable value of **a**. In Loglan **matmymatma**, *maternal grandmother*, the two vowels occurring after *tm* should sound distinctly different.

The vowels have two series of names, the legacy VCV forms with the shape V-**fi** (V-**ma** for capitals) and the newer series with the shape **zi**-V. Another series of legacy vowel names with the shape V-**zi** is associated with the Greek alphabet. The legacy vowel names are phonetically weird among Loglan words, but they seem to be reasonably well accommodated. It should be noted that the principal function of all “letteral” words in Loglan is not as names of letters, but as *pronouns*, as will be discussed subsequently.

²A suggestion of John Cowan

³It is *not* easy to insert this in LaTeX!

2.1.2 Consonants

The Loglan consonants are **b,d,f,g,h,j,k,l,m,n,p,r,s,t,v,z**. The pronunciations of these are standard European pronunciations, except that **c** is English *sh* and **j** is the corresponding voiced sound found in English *azure*. **g** is always “hard”. **h** has an alternative pronunciation as *ch* in Scottish English *loch* when final in a syllable⁴ [this is NEW: **h** does not occur in syllable final position in 1989 Loglan]. **n** is pronounced as *ng* in English *sing* when it appears before **k,g** or syllable final **h**. Note that Loglan **ng** is always as in English *finger*. NEW: **q,w,x** with odd pronunciations are eliminated: this was already mostly carried out in the 1990’s. It is worth noting specifically that **tc** is the sound in English *church* and **dj** is the sound in English *judge* (and that Loglan **j** is *not* the latter sound).

The names of the consonants are of the forms C-**ai** (capitals) and C-**ei** (lowercase), with a third series C-**eo** associated with the Greek alphabet. It should be noted that the principal function of all “letteral” words in Loglan is not as names of letters, but as *pronouns*, as will be discussed subsequently.

The Loglan alphabet is **abcdefghijklmnpqrstuvyz**, the Latin alphabet without **qwx**. The foreign letters **qwx** have proposed names **Kaiu, Keiu, Vaiu, Veiu, Haiu, Heiu** as part of a general proposal reserving C-**aiu**, C-**eiu** as additional series of letterals. Names for these letters are important, as all have common use as mathematical variables, the prototype for the Loglan use of letter names as pronouns.

2.1.3 Vowels pretending to be consonants

The vowels **i** and **u** are sometimes pronounced as the English consonants *y* and *w*.

2.1.4 Consonants pretending to be vowels

The continuants **m,n,l,r** can appear as syllabic consonants (functioning as the vowel in a syllable). In this role, these consonants are doubled, **mm, nn, ll, rr**. The requirement that syllabic consonants be doubled is NEW, but it is actually suggested by JCB in Loglan 1.

⁴I accept an amendment from John Cowan allowing the alternative pronunciation of **h** in all contexts.

A doubled continuant may not appear adjacent to another occurrence of the same consonant in the same word (i.e., without an intervening pause in speech).

2.2 Diphthongs and vowel grouping

There are some mandatory and some optional pairs of vowels which form diphthongs⁵, which can serve as the vowel component of a single syllable.

2.2.1 Mandatory diphthongs

The mandatory diphthongs are **ai**, **ao**, **ei**, **oi**. The pronunciations are as one would expect from the values of the vowels, except that **ao** is as in English *cow*.

A new proposal is the addition of two irregular mandatory diphthongs **iy** and **uy**, allowed only in *cmapua*: the practical use of this for the moment is that it makes the new-style name **ziy** of the letter **y** legal. These are pronounced *yuh* and *wuh*, not what the English-reading eye expects.

2.2.2 Optional diphthongs

The optional diphthongs are the pairs of vowels beginning with **i** or **u**. The pronunciation of these diphthongs is as if the initial **i** were the English consonant *y* or the initial **u** were the English consonant *w*.

When a pair of adjacent vowels is not pronounced as a diphthong (i.e., when there is a syllable break between them), one may flow into the other without pause, or a glottal stop (not expressed in writing except possibly indirectly by a hyphen or stress mark) may be inserted (NEW: we do not allow the glottal stop to be an allophone of the pause, as earlier versions of Loglan did; in Lojban they require insertion of what we would write **h** in non-diphthong vowel pairs⁶). It is noted in our sources that such a pair is

⁵I looked this word up in the dictionary, and it does seem to refer to vowel combinations in a single syllable, which is what is desired.

⁶We note that the orthography **la'e** for a typical disyllabic CVV *cmapua* in Lojban, which they read **lahe**, is also valid orthography for this word in TLI Loglan, with the apostrophe signalling stress on the first syllable. We note without recommending it that a dialect in which non-diphthongs are broken with an invisible **h** is perfectly possible, as long as the additional rule is made that the **hV** *cmapua* must be pronounced with the al-

easier to pronounce if one of the syllables is stressed and the other is not, and when the two vowels are the same this is required.

2.2.3 Expanded notes on diphthong pronunciation

Two-letter diphthongs pronounced monosyllabically are as follows:

ai is English long **i** as in *pine*

ei is English long **a** as in *pane*

oi is as in English *boil*

ao is as **ow** in English *cow* (this is an irregularity, but we are stuck with it).

These four are the mandatory monosyllables: where these letters are grouped together, they must be pronounced monosyllabically.

The pairs **ia**, **ie**, **ii**, **io**, **iu** are optional monosyllables. They may be pronounced as two syllables (smoothly moving from one vowel to the other without pause) or monosyllabically by pronouncing the initial **i** with the usual consonantal value of English **y**.

The pairs **ua**, **ue**, **ui**, **uo**, **uu** are optional monosyllables. They may be pronounced as two syllables (smoothly moving from one vowel to the other without pause) or monosyllabically by pronouncing the initial **u** with the usual consonantal value of English **w**.

These two classes are all the optional monosyllables. The disyllable pronunciation may be forced by an explicit syllable break (one of **-’***); some contexts without an explicit marker force the monosyllabic pronunciation, but I believe that no context forces the disyllable pronunciation in the absence of an explicit syllable break.⁷

The other disyllables are obligatory disyllables: they should be pronounced with a smooth movement from one vowel to the other without pause. Pronunciation is assisted if one is stressed and one is not. In the repeated vowel disyllables **aa**, **ee**, **oo**, one of the syllables must be stressed and the other must be unstressed. An explicit stress marker is permitted to indicate which one is to be stressed, but is certainly not required. The same stress

ternative pronunciation of **h** (using the alternative pronunciation with **hVV** cmapua would be optional). Of course the invisible uses of **h** would have the usual soft pronunciation of **h**.

⁷There *were* such contexts in the past: a **CCVV** or **CCCVV** predicate word with the **VV** an optional disyllable had of course to be two syllables. But both these shapes for predicate words are now banned.

rule applies to **ii** and **uu** where these are pronounced disyllabically. I am open to the idea of a pronunciation of disyllables using a glottal stop, as I do not regard a glottal stop as an adequate implementation of Loglan pauses.

We summarize some refinements. The **i**-final mandatory diphthongs are not read as such if immediately followed by another **i**: e.g., **aii** groups as **a-ii**. The triples **iuu** and **uii** are grouped **i-uu** and **u-ii**. All the preceding rules in this paragraph are designed to avoid formation of **i-i** or **u-u**, which would trigger the doubled vowel stress rule. The triples **iii** and **uuu** are recognized as triggering the double vowel rule (though the orthography does not tell us how).⁸ The odd triple **aoa** does group as **ao-o** and also triggers the double vowel rule. The triple **aoi** is grouped **ao-i** and is not regarded as containing a broken mandatory monosyllable.

2.2.4 Grouping long streams of vowels in predicate words and names

A string of three or more vowels in a name or predicate word which is not marked with explicit syllable breaks (a hyphen or a stress marker is used for this purpose, as we will discuss below) is resolved into syllables (possibly including consonants adjacent to the string of vowels) following a priority order reading left to right

1. group the first two vowels and continue if they make up a mandatory diphthong (or **ii**, **uu**). Do not apply this rule to a mandatory diphthong ending in **i** immediately followed by another **i**.
2. pronounce the first vowel as a single syllable and continue if the second two make up a mandatory diphthong (or **ii**, **uu**) and the first two do not (or in the cases **aii**, **eii**, **oii**).
3. optionally, take the first vowel as a syllable and continue or group the first two vowels and continue if they make up an optional diphthong and the previous conditions do not hold; the parser will always take the second alternative.
4. pronounce the first vowel as a single syllable and continue, if none of the previous conditions hold.

⁸I have adopted the view that **i-iV** and **u-uV** are in general impossible to distinguish from **i-V** or **u-V** and should simply be forbidden, implemented 5/11/18.

5. by “continue”, we mean “apply the same set of rules to the remainder of the stream of vowels”.

This is NEW, not the same as the rule in 1989 Loglan (given there only for names) but it appears to have similar effects in practice.⁹

2.2.5 Grouping long streams of vowels in structure words

A stream of vowels of even length parsed as a *cmapua* (structure word) will first be divided into VV units, each of which will then be read as one or two syllables depending on whether it is or can be a diphthong. A stream of vowels of odd length so parsed will not occur, as it would be read as a V word followed by a stream of VV units, and a pause is required before the first of the VV units (see below) (this would equally be the case if it were read as a stream of VV units followed by a V word: it would be necessary to pause before the V word).

2.2.6 Vowel pairs with optional grouping revisited

An optional diphthong not appearing in a stream of three or more vowels may be pronounced either as a single syllable or two syllables: sometimes other factors will force the monosyllabic pronunciation. We believe that it is not possible to force the disyllabic pronunciation of an optional diphthong without explicit indication of a syllable break.

2.2.7 Doubled vowels and stress

Where doubled vowels are not separated by a pause and not pronounced as a diphthong, one of them must be stressed: this always applies to **aa**, **ee**, **oo**, and applies to **ii**, **uu** unless they are pronounced *yee*, *woo*.

⁹The rule in 1989 Loglan calls for right-grouping, which is psychologically incredible. I think the reasons why JCB wanted to right-group are captured by the second clause of my algorithm.

2.3 The Loglan syllable

2.3.1 Discussion of stress and notation for syllable breaks and degrees of stress

We now discuss the Loglan syllable. Each Loglan syllable is either unstressed, stressed, or emphatically stressed. Syllables may be separated by a hyphen. A marker ' of stress or * of emphatic stress may terminate a syllable¹⁰: a stress marker may not be followed by a hyphen, as the stress marker itself serves the purpose of separating the syllable from a following syllable (though a stress marker can also be final in a word; actually, so can a hyphen, in a phonetic transcript where no explicit pause occurs between a word and the following word). Syllable breaks and stress markers do not have to be written explicitly, though it can be useful to do this. The precise definition of the syllable that we give does not appear in the sources, but every component of it is found there, and all words in the dictionary are successfully resolved using this definition; we do not regard this definition as new except in detail. The use of the hyphen and the explicit stress markers is NEW. The hyphens and stress markers make possible a style of writing Loglan with no whitespace except where explicit comma-marked pauses occur, which we refer to as “phonetic transcript”. The availability of phonetic transcript means that we have our own “native” phonetic notation and do not need to appeal to IPA notation or to such odd expedients as JCB uses.

Note that a syllable without a stress marker is not necessarily unstressed: we simply have not committed ourselves.¹¹

We briefly remark on the capitalization conventions of Loglan. A stream of letters and junctures is governed by the capitalization convention. The basic rule is that a lower-case letter will not be followed directly by an upper-case letter (except that **z** may be followed by a capitalized vowel (useful in acronyms, as **DaiNaizA**), and that embedded copies of letter names may freely be capitalized (as in **leSai** and as also seen in **DaiNaizA**). Note that capitalization may resume at a juncture, as in **Beibi-Djein**. Note also that this rule allows all-caps.

We briefly note that stress in Loglan names is completely free, as is

¹⁰It is important to note that it is the end of the syllable, not the vocalic unit of the syllable, which is marked with the stress.

¹¹Many syllables in actual Loglan text can be recognized as definitely unstressed from context, e.g., most syllables in predicate words, but there is no marker for this so far.

stress in Loglan structure words (cmapua): in words of these classes one may have no stressed syllables, one stressed syllable in any desired position, or no stressed syllables. In Loglan predicate words, there is a single primary stress on the penultimate syllable (usually; sometimes a special additional unstressed syllable intervenes, governed by the rule that if the actual penultimate syllable has a syllabic consonant or **y** as its vocalic unit, it cannot be stressed, and in this case the preceding syllable is stressed); there may be additional stresses on the penultimate syllables in *borrowing djifoa* components of predicate words, as described below. There is a special rule that a pause must intervene between a finally stressed structure word and a following predicate word. Explicit stress markers may be used to signal an unexpected stress in a foreign name, or may be used by the Loglan writer as a mechanism for rhetorical emphasis in any word or to signal actual stress in reported speech.

2.3.2 Every syllable has a vocalic unit

Each Loglan syllable contains a vocalic unit, which is either a single vowel¹², a diphthong pronounced as such, or a doubled continuant.

2.3.3 Like Gaul, every syllable has (up to) three parts

Each Loglan syllable consists of up to three parts.

1. The first part (which is optional) is a consonant cluster called the initial consonant group. There is a list of pairs of consonants called permissible initial pairs (or just initial pairs). An initial consonant group will be either a single consonant, an initial pair, or a group of three consonants in which each adjacent pair of consonants is a permissible initial pair.¹³
2. The second part is the mandatory vocalic unit.

¹²the irregular vowel **y** can be a vocalic unit, though its distribution is limited; it can occur freely in names and as a “phonetic hyphen” in complex predicate words, but not at all in borrowed predicate words, nor in cmapua except in the so far very rare diphthongs **iy**, **uy**.

¹³The initial pairs are **bl br ck cl cm cn cp cr ct dj dr dz fl fr gl gr jm kl kr mr pl pr sk sl sm sn sp sr st sv tc tr ts vl vr zb zl zv**

3. The third part, which is optional, consists of one or two final consonants. There is a list of impermissible medial pairs and a list of impermissible medial triples.¹⁴ A final consonant may not be followed by a vowel and may not stand at the beginning of an impermissible medial pair or triple (regardless of whether the other consonants are in the same syllable, and ignoring stress marks and hyphens). A syllabic pair is an impermissible medial pair: a final consonant may not participate in such a pair, even with a syllable break intervening.

A pair of final consonants may not consist of a non-continuant followed by a continuant [the last sentence is *NEW*, but seems self-evident: such a pair would basically have to be pronounced as a separate syllable, and no violations occur in the dictionary].

In placing syllable breaks in the absence of an explicit hyphen or stress mark, the rule is followed that a final consonant shall not stand at the beginning of a legal syllable, except in the case of a first final consonant at the end of a syllable whose initial consonant group consists of a single consonant and whose vocalic group is a single regular vowel (the general idea is that the syllable break is usually placed as early as possible, except that a CVC syllable is read in preference to a CV syllable where possible).

2.3.4 Words must resolve into syllables

Every Loglan word must resolve into syllables. Some classes of words must also resolve into other small units which are not exactly syllables, though syllables do not as a rule cross their boundaries.

Syllable boundaries may be phonemic (in the sense that changing their placement can change a word into a different word) only in the case of syllable breaks between vowels, and only in names. Potential words which can be written only with the use of explicit syllable breaks may be legal names (as the classic example **La Lo-is**) but not legal words of any other class.

¹⁴The impermissible medial pairs consist of all doubled consonants, any pair beginning with **h**, any pair both of which are taken from **cjsz**, **fv**, **kg**, **pb**, **td**, any of **fkpt** followed by either of **jz**, **bj**, and **sb**.

There is a list of impermissible medial triples as well, consisting of **cdz**, **cvl**, **ndj**, **ndz**, **dcm**, **dct**, **dts**, **pdz**, **gts**, **gzb**, **svl**, **jdj**, **jtc**, **jts**, **jvr**, **tv1**, **kdz**, **vts**, and **mzb**. All of these consist of a consonant followed by an initial pair, but they are not permitted to occur with the juncture between syllables in either of the two positions.

2.3.5 Notation for consonant and vowel patterns

C represents a consonant; cc (lowercase pair) represents an initial pair, where CC represents any pair (initial pairs included). V represents a regular vowel (not **y**); vv (lower case) represents a diphthong pronounced as such, where VV represents any pair of regular vowels, whether pronounced as a diphthong or not. In such pattern notations, a hyphen stands for an explicitly marked syllable break, whether marked with a hyphen or a stress marker.

2.4 Word Forms

2.4.1 Word forms enumerated

There are four phonetic classes of Loglan words: these are (1) structure words (**cmapura**), (2) name words, (3) borrowed predicate words and (4) complex predicate words.¹⁵

2.4.2 Pauses and word boundaries

Words (in the phonetic sense) end at whitespace, at a comma or mark of terminal punctuation (periods and some other punctuation marks), or sometimes without any explicit indication at all (where phonetics are sufficient to recognize where one word ends and another begins).¹⁶

A comma in Loglan marks an explicit pause (and is followed by whitespace; the close comma used to indicate unusual syllable breaks between vowels in earlier versions of Loglan is replaced by the hyphen which we use to represent syllable breaks in general).

¹⁵The phrases “complex predicate word” and “borrowed predicate word” may on occasion be shortened to “complex predicate” and “borrowed predicate”, or even “complex” and “borrowing”. We comment on this because a “predicate” proper is actually a grammatical constructon.

¹⁶The marks of terminal punctuation are **::;?!.** The silence or change of voice marker **#** used by JCB is supported. This may not appear in quoted or parenthesized Loglan text; it is not really fully privileged punctuation. It does allow multiple utterances in different voices (including the same voice stopping and starting again) on the same line of parsed text. Words can end at double quotes or close parentheses, which appear in certain special contexts. The dash **–** and ellipses **...** are supported as “free modifiers” and may be used fairly freely.

Whitespace is sometimes an explicit pause and sometimes a word boundary which is not marked by any actual phonetic feature. Where whitespace does not appear, one should not pause. Where a pause is allowed at whitespace, a comma should always be permitted (The old parser LIP does not always support this, but we regard this as debugging, not a novelty). There are situations where whitespace is allowed due to a word break but an actual comma pause would change the parse (and so in speech such a whitespace is not expressible as a pause).¹⁷

Vowel initial words are always preceded by a pause if they are not at the start of a text or utterance. Consonant final words are always followed by a pause if they are not at the end of a text or utterance. Thus, whitespace preceded by a consonant or whitespace followed by a vowel must represent an actual pause. So we also regard whitespace preceded by consonants or followed by vowels as an explicit pause. It appears to be NEW that we must pause before the first in a stream of VV words, but it is also clearly necessary, as experiments with phonetic transcripts have revealed.¹⁸

We note the subtle point that the end of a predicate word may have to be indicated by whitespace if the stressed syllable is not explicitly marked. So in this case whitespace may have no local phonetic meaning but will have the definite phonetic effect of signalling the presence of an earlier stressed syllable. In phonetic transcripts, where whitespace not representing pauses is suppressed, the stresses in predicate words must usually be marked explicitly.

It is possible to resolve a stream of Loglan phonemes into words unambiguously, with the qualification that the resolution of streams of Loglan grammatical particles is actually done by the grammar proper. We indicate how to do this.

2.4.3 Structure words (**cmapua**)

The structure words or “little words” (Loglan **cmapua**) are the grammatical vocabulary of Loglan, for the most part.

¹⁷There are such instances of whitespace which are permitted to be written but cannot represent a pause, in connection with the handling of the legacy forms of the APA connectives.

¹⁸The new PEG implementation supports the traditional requirements that pauses at the end of a serial name and pauses before a logical connective must be actual comma pauses. Some logical connectives are consonant-initial: there is a purely phonetic description of the front of a logical connective in the PEG.

A structure word (in the phonetic sense: some phonetic structure words are actually semantically names or predicate words) is a word which resolves into a stream of V, VV, CV, CVV, and Cvv-V units (where vv denotes a diphthong). These words are called **cmapua** in Loglan. The CVV units do not have to be syllables (there is no requirement that the VV be a diphthong, or that it be pronounced as such if it is an optional diphthong). A V unit can only appear initially; if any unit in a cmapua is of the shape VV, all units are of the shape VV. We recall the rule that one must pause before a word which begins with a vowel, so one must pause before an initial V- unit; it is not necessary to pause between VV units in a phonetic word made up of such units, but it is necessary to pause before a lone VV unit or the first in a stream of such units (the necessity of pausing before the first in a stream of VV units seems not to have been recognized in NB3). Except in the case of the end of a string of VV units, it is impossible to recognize the ends of individual words in a stream of cmapua phonetic units on phonetic grounds alone: the grammar proper allows us to resolve streams of unit cmapua into words, but for phonetic purposes we may regard streams of VV units and streams of non-VV units as “words”.¹⁹ To support names for the letter **y** (and perhaps add a little cmapua space), we propose to allow the irregular monosyllabic diphthongs **iy** and **uy** in CVV (but not Cvv-V) cmapua units. We further propose to allow **y** to be a V djifoa unit (in compounds). This makes the name **yfi** of **y** legal.

It is part of the definition of a cmapua unit that it cannot be an initial segment of a predicate (either complex or borrowed): a more precise statement of this is that a cmapua unit may not be followed without an intervening explicit pause by **y** or by **CyC** (nor may a CV cmapua unit be followed by **CCy**), and may only be followed without an intervening explicit pause by **CC** if the **CC** itself is initial in a predicate word (this last condition requiring lookahead!) This indicates the (sometimes rather subtle) way that the end of a phonetic structure word at the beginning of a predicate word is recognized; how to recognize the end of a structure word at the beginning of a name word is covered in the next section (and may require considerable lookahead).

There is a precise set of conditions under which an apparent cmapua unit cannot be part of a cmapua word because it is the start of a predicate word

¹⁹The parser will end a stream of cmapua units before a cmapua unit which is either a name marker, a marker of an alien text construction, or an initial marker of a quotation construction; a stream of cmapua units can begin with a name marker word if it does not start a name word.

(if it is anything legal). It is further the case that the start of a borrowing or complex not beginning with an initial pair or triple of consonants must pass one of these tests:

1. If the unit is followed without pause by two or more consonants, where the consonant group is not an initial group of consonants or is a consonant or initial group of consonants followed by a syllabic pair.
2. If the unit stands at the beginning of a copy of a *CVVy*, *CVCy*, or *CVCCy* djifoa (see below for discussion of djifoa).
3. If the unit is finally stressed and immediately followed without pause by more than one consonant, except in the case where the unit is a *V* or *Cvv-V* *cmapua* unit or the head of a stream of *VV* *cmapua* units followed by *ccV*: this last cluster of cases is simply illegal, as no predicate word can start in such a way for technical reasons. An actual *V* or *Cvv-V* *cmapua* unit may be followed without pause by *ccV* (the *cc* being an initial pair starting a predicate word) if not finally stressed.
4. If the unit is followed by an initial group of consonants which is then followed by *V* or *VV* followed by whitespace or non-syllable-break punctuation or by a stressed *V* followed by a single *V* not in a diphthong.

We note that the parser applies these exact tests rather than looking ahead to see if a predicate word actually starts at the beginning of the apparent unit under consideration.

Where the final unit of a phonetic *cmapua* immediately precedes a predicate word and is stressed, it must be followed by a pause: if the stress is marked in writing and the following predicate word is consonant initial, the pause must be explicitly written as a comma. We are pleased with the fact that this rule, going back to the beginnings of Loglan, can be expressed in our orthography and is enforced by our parser; it is quite invisible to LIP.²⁰

²⁰We note a subtle point about the articulation of acronyms in Loglan: these are semantically names but phonetically *cmapua*. The legacy vowel names are of the weird shape *VCV*; they can occur without initial pause following a *CVV* unit because the *(CVV)(VCV)* shape (when articulated as letters) can be rearticulated as *(Cvv-V)(CV)* for purposes of articulation as *cmapua* units. We have abandoned an irregularity found in the previous provisional parser: the *CVV* letters of the common sorts have the *VV* actually a diphthong, but there are *Ceo* letters, and these formerly worked in acronyms preceding a legacy vowel, but no longer do.

2.4.4 Name words and the false name marker problem

Semantically, name words are *proper names*, as their name suggests. Name words are usually capitalized, but this is not obligatory. Names are required to end with consonants. A Loglan name borrowed from a foreign source which happens to end in a vowel is conventionally terminated with an **s**, as in **Selis**, *Sally*. Names borrowed from foreign sources will be spelled in such a way that the Loglan speaker's pronunciation of them will approximate the original (**Ainctain**, not **Einstein**). There is a direct foreign name construction which allows preservation of foreign spelling, to be discussed later. Names derived from Loglan words which are not names (and so vowel-final) are conventionally constructed by adding an **n** to the Loglan word.

A name word is a stream of Loglan syllables ending in a consonant followed by whitespace, a comma marked pause, or terminal punctuation. Name words are the only Loglan words which end in consonants. The parser currently requires an explicit comma pause in place of whitespace after a name word in most contexts, though the fact that a consonant followed by whitespace is recognized as an explicit pause suggests that this can be relaxed. The boundary of a name word on the right is readily recognized (the consonant followed by a explicit pause or terminal punctuation). Left boundaries of name words always fall just after either a *cmapua* belonging to a class of "name markers"²¹, or an explicit pause. A candidate left boundary for a name word falls just after an explicit pause or just after a phonetic copy of a name marker such that the text between its end and the recognizable right boundary of the name word resolves into syllables (and of course contains no pauses or whitespace). A candidate left boundary is said to be marked if it is just after a name marker or just after an explicit pause immediately preceded by a name marker. The actual left boundary of a name word is the leftmost marked candidate left boundary, if there is one, and otherwise the rightmost candidate left boundary.

We refer to name markers which are candidate left boundaries of name words but are not the actual left boundaries as "false name markers". Earlier versions of Loglan forbade these, with the odd effect that (for example) **la**

²¹The name markers are **la**, **hoi**, **hue**, **ci**, **liu**, which must be name markers for one reason or another, and **gao** if a proposal to allow this *cmapua* to form letter names from name words is accepted, and **mue** in support of a proposal to allow name words to be used as dimensions. A proposal of mine that the words of social lubrication **loi**, **loa**, **sia**, **sie**, **siu** be name markers has been withdrawn.

could not occur in a name. Lojban still forbids this (and has very few name marker words). Our current grammar of Loglan strongly restricts the situations in which any name word can occur with an unmarked left boundary (for example, the current Academy has banned unmarked vocatives), and in any such context there is provision for the boundary to be marked with a suitable name marker if the name word happens to contain a false name marker. This is a generalization of rules for handling serial names adopted in the 1990's.

Where a false name marker occurs and it was the actual intention of the speaker to make it the left boundary of the name word, the intention can be realized by pausing explicitly after the name marker, making it a true name marker. The intention would more usually be realized by pausing somewhere earlier, due to the actual class of errors which leads to this situation, discussed in the next paragraph.

The parser will raise an error if it finds a name marker word followed possibly by an explicit pause, followed by text including whitespace but no explicit pause, followed by a name word (recognizable by its right boundary at a consonant). The problem with this is that because of the intervening whitespace, it cannot be the intention of the writer that the name marker sets the left boundary of the name word, but if the utterance were read without pause, the name marker would indeed set the left boundary of the name word.²² There are uses of name markers in which they are not followed immediately by name words: it is the obligation of the speaker to explicitly pause at some point after such an occurrence of a name marker and before the next occurrence of an actual name word, and Loglan orthography requires this to be indicated explicitly. Note that whitespace before a vowel or after a consonant does suffice, but where this doesn't happen, an explicit comma pause (of the form V, C) may be required. Actually complying with this rule is best implemented by style directives such as “always pause after a predicate name”, rather than by attention to this esoteric rule as such.²³

²²A reader requested an example. **La Farfu ga cluva la Djan** would trigger this problem, and needs to be corrected to **La Farfu, ga cluva la Djan**. Without the explicit pause, it could be read phonetically as **lafar'fuclu'valadjan**, which parses as **La Farfugacluvaladjan**, a single name.

²³A directive which would always work is, “always pause after the first word after a name marker, whether it is a name word or not (and whether you paused after the name marker or not)”. When what follows the name marker is not a name word, one need not pause after exactly one word, but one should pause at some natural point before the next

The rule that false name markers cannot occur in names has already been abandoned (in the 1990s) by TLI Loglan; it still obtains in Lojban, which has very few name markers. Precise definition of what you do with false name markers was hard to think about before phonetic parsing was available. The requirement that names resolve into syllables is NEW (I seem to recall that Lojban does something like this?), and interacts with the definition of a false name marker as indicated above. The automatic detection of dangerous situations where a non-explicit pause should be made explicit is NEW, and we can report from extensive experience in parsing Alex Leith’s Visit to Loglandia that there are straightforward ways to correct problems it detects (and that it really does detect things which are potential problems). We can also report, based on extensive parsing of phonetic transcripts of Loglan utterances, that the rules stated above appear to work: the situations in which a name is unintentionally started too early can be controlled.

It might be thought that imposing additional restrictions on the formation of Loglan names would damage the corpus. In fact, there are only two difficulties which arise. Continuant must be doubled (as in **Rrl**, “Earl”) and this happens not unseldom; but JCB *did* actually suggest that this might be a good idea in Loglan 1. The other problem is that our definition of the syllable does not allow final consonant clusters of more than two consonants. Usually one of the consonants in such a cluster is a continuant, so this can be fixed, as in **Hollmz**, **Marrks**.

2.4.5 Borrowed predicates

Semantically, borrowed predicates are common nouns, verbs, adjectives or adverbs (not names) borrowed from foreign sources. Their phonetics are dictated by the idea that they should not take any of the phonetic shapes of “native” Loglan predicate words (complexes) other than the primitive five-letter shapes CCVCV or CVCCV (in which case they are phonetically treated as (very simple) complexes). There will be a discussion later of how to modify the foreign originals to get legal Loglan predicate words which are not phonetically of the shape of a complex.

Phonetically, all predicates (borrowings and complexes) have certain common traits. They are *penultimately stressed* (with the exception that a special

name word. An example: in **hoi le farfu je la Rabrrt**, which is bad, it is legal but odd to amend it to **hoi le, farfu je la Rabbrrt**, most natural to amend it to **hoi le farfu, je la Rabrrt**, and a sign of last minute panic to amend it to **hoi le farfu je la, Rabrrt**.

unstressed syllable with a continuant or **y** as their vocalic unit may intervene between the stressed syllable and the final syllable). They are *vowel-final*. They contain a *consonant pair*, either CC or CyC (the latter in some complexes).

A borrowed predicate is a stream of Loglan syllables which contains just one stressed syllable, whose vocalic unit is not a continuant pair, the stressed syllable being followed by at least one and no more than two syllables, the first of these, if there are two, having a continuant vocalic unit, and the final one not having a continuant vocalic unit and not ending in a consonant: in other words, it is penultimately stressed, ignoring one possible intervening unstressed continuant syllable, and ends in a vowel. A borrowed predicate contains at least one pair of distinct adjacent consonants; it is permissible for one of these to be in a continuant pair. A borrowed predicate may not contain two successive syllables with continuant pairs nor may it start with such a syllable (or, as noted, end with such a syllable), nor may a continuant pair immediately follow a vowel in a borrowed predicate. A borrowed predicate may not contain a doubled vowel unless the doubled vowel is pronounced as a diphthong. A borrowed predicate may not contain **y**. The part of the borrowed predicate before the first pair of distinct adjacent consonants must have the property that omitting it (and dropping any explicit syllable break between the pair of consonants) will not leave a legal borrowed predicate. A borrowing cannot begin VccV with the cc permissible initial, and there can be no ccVV or cccVV borrowed predicates²⁴. A CvV-V cmapua unit or a stream of VV cmapua units followed by ccV cannot be the shape of a borrowing: this restriction is caused by a technical problem with borrowing djifoa (see below). The last restriction is NEW, motivated by the same problem averted by forbidding VccV-initial borrowings; only one word in the dictionary was affected.²⁵

²⁴The ccVV predicate words are forbidden so that CVCccV complex predicates do not have to be **y**-hyphenated; the cccVV predicate words would, if allowed, greatly complicate our parsing algorithm for a weird technical reason.

²⁵What needs to be averted is the possibility of reading a borrowing djifoa ending with ccV**y** as a stream of phonetic cmapua units followed by a predicate word beginning with the ccV**y** djifoa instead of the intended borrowing djifoa. Both the original VccV rule and my new rule forbid some predicate word shapes which are actually not problematic (but also not needed). I have modified this proposal recently to allow more of the improbable space of possible borrowings with many vowels before the CC pair to be preserved; these very likely should be banned, but this rule is not the context in which the decision should be made. I note the possible reasonable restriction that no predicate word should begin

The beginning of a borrowed predicate must either be at an initial pair or triple of consonants or a CV^n passing one of the tests listed above in the *cmapua* section (and thereby failing to be a *cmapua* unit). The end of a borrowed predicate is recognized either by seeing an explicitly stressed syllable and counting the one or two allowed following unstressed syllables, or by whitespace, a comma, end of text, or terminal punctuation: in the latter case, the fact that one is in a borrowed predicate is recognized by the occurrence of two adjacent distinct consonants.

A borrowing must not resolve into *djifoa* (see the next section), even ones with badly placed internal syllable boundaries or lacking required phonetic hyphens (noting that the five-letter *djifoa* and the borrowing *djifoa*, when not final, *include* their **y** hyphens and so will not be involved in such resolutions, since a borrowing candidate cannot include **y**, and a final borrowing *djifoa* is always preceded by **y**). This both ensures that actual complex predicates are read as complex predicates rather than borrowings, and ensures that certain illegal complexes are not read as legal borrowed predicates.

There is almost nothing actually new in the description of borrowed predicates (only the one point labelled with NEW above): some features are points worked out in the 1990's (all details of borrowing *djifoa* are late and not in 1989 Loglan 1: these details motivated the elimination of **y** from borrowings and the more baroque excluded forms above). Forbidding doubled vowels in borrowings is the most recent change, made by the current Academy in the last few years. Everything else is explicit in the sources somewhere (there may be some guesswork about the exact rules for use of continuants for gluing, but they fit actual practice). The precise definition of the syllable was made in order to make it possible to implement the description of borrowings in NB3 and L1.

2.4.6 Complex predicates

Semantically, the complex predicates are the native common nouns, verbs, adjectives and adverbs of Loglan (which really does not have these parts of speech). The root vocabulary of Loglan, apart from the *cmapua* which make up its grammatical vocabulary, is made up of primitive predicate words of the five-letter forms CCVCV and CVCCV and complex predicates built from these five letter roots and shorter combining forms (the deprecated original

CV⁴ or V⁴.

terminology for these is “affixes”; we now call them *djifoa*, a Loglan term) of shapes listed below: each five letter root has a form in which its final vowel is replaced with **y** which may appear in non-initial position in a complex, and may have one or two three-letter forms of the shapes CCV, CVV, CVC associated with it. Here we concern ourselves with the phonetic rules by which such complex predicates are to be constructed. Recall that Loglan predicate words have the common characteristics of *penultimate stress*, being *vowel-final*, and having a *consonant pair*.

A complex predicate is a stream of units distinct from syllables, called *djifoa*, with the additional property that any syllable breaks respect the *djifoa* boundaries (by which we mean that no syllable contains parts of two neighboring *djifoa*; testing for resolution into *djifoa* requires that *djifoa* with badly placed internal syllable boundaries be recognized; predicate words which resolve into *djifoa* with badly placed boundaries are to be rejected as borrowings as well).

The *djifoa* take the forms

1. CVV (legal syllable forms Cvv or CVV or CV-V)
2. CVC
3. ccV (the cc must be permissible initial).
4. ccVCV (when this is not final, the final V is replaced with **y**). The cc must be permissible initial. The only legal syllable break is ccV-CV.
5. CVCCV (when this is not final, the final V is replaced with **y**). The CC must not be impermissible medial. CV-ccV (if the CC is permissible initial) and CVC-CV are legal syllable breaks. Either break is permitted if the CC is permissible initial.
6. a borrowing predicate (modified to have final rather than penultimate stress if not in final position) with appended **y** (in an unstressed syllable by itself) if not in final position.
7. When attempting to resolve a predicate word into *djifoa* to establish that it cannot be a borrowed predicate, the forms to consider are CVV, CVV**r**, CVV**n** (only when followed by **r**), CVC, ccV, and the five-letter forms in final position only, plus the illegal CV-C, c-cV, c-cVCV (the last in final position, the hyphen indicating an explicit syllable break or

stress marker; c-c denotes an initial pair broken by an explicit syllable break). A mechanical resolution into these forms, without any side conditions, shows that a string cannot be a borrowed predicate. We do not need to list forms with syllable breaks C-V because these are explicitly forbidden by our phonetics already.

Please note that we are not saying anything about where djifoa come from, not because this is unimportant, but because it does not bear on the strictly phonetic business at hand.²⁶

Note that a complex predicate may not contain any continuant pair, except in the context of a borrowing djifoa.

Any complex includes at least two djifoa, unless it consists of a single five-letter djifoa (the latter could also be viewed as a separate species: primitive predicates).²⁷

Djifoa appearing in non-final position may have phonetic hyphens appended, which may take the shapes **n**, **r**, or **y**. Only one phonetic hyphen can be appended to any djifoa. The consonant hyphens can only be appended to CVV *cmapua*. A phonetic hyphen **n** appears only when followed by **r** in the next djifoa. A phonetic hyphen is never final in a complex predicate nor will it follow a five-letter djifoa. A phonetic hyphen **y** is always unstressed, and appears by itself in a syllable or in a **Cy** syllable in the context CV-**Cy** of a CVC djifoa (CVC-**y** also being allowed; the **y** in a five-letter djifoa in non-final position may participate in a **Cy** or **ccy** syllable, which is also always unstressed). A borrowing djifoa is always preceded by **y** if not initial (the **y** will be a hyphen or a constituent of a five-letter or borrowing djifoa) and includes an appended **y** if not final.

An initial CVV followed by a djifoa beginning CV must be hyphenated with a consonant. An initial CVC followed by C in a way which would make a permissible initial pair must be hyphenated with **y** if the entire word is not of length 6. These are rules to allow recognition of the start of a predicate word, preventing *cmapua* units from falling off the front.²⁸

²⁶This goes along with our not explaining where the primitive predicate words of the five-letter forms come from: this is fundamental data of the language which is not going to change. I do not expect that more djifoa will be created from five-letter predicate words, so I do not need to explain how it was done. The curious may look in older sources; this has no bearing on learning the language.

²⁷The PEG grammar also thinks that a borrowing is a one-complex djifoa, so it classes all predicate words as complexes!

²⁸No additional enforcement of this is needed in the parser: the algorithm described

There must be a CC pair (or a CyC pair) in a complex predicate. This is enforced by two provisions: the start of a complex must be at a CC pair or at a CVⁿ passing one of the tests listed in the cmapua section (and thereby failing to be a cmapua unit); further, any CVVy djifoa must be followed by a complete complex or borrowing (a complex made entirely of CVV's using y-hyphens would not contain a CC or CyC pair, and is excluded by this rule). This last condition is NEW: LIP accepts **ceaydea** as a predicate word, for example).

A complex predicate must end with a regular vowel (so the last djifoa will not be CVC or phonetically hyphenated). Any adjacent pair or triple of consonants in a complex predicate may not be impermissible medial.

A complex predicate may contain stress only in the penultimate position (where it must be stressed; in determining the penultimate stress, a syllable with y intervening between the stressed syllable and the final syllable may be ignored) or in the final position of a non-final borrowing djifoa (just before the y; the final, not the penultimate syllable, of the borrowing; here stress is optional, unless such a syllable is also penultimate in the predicate word). It is also permitted to contain a pause (in explicit comma form) after the y at the end of a stressed non-final borrowing djifoa (violating our expectations about word boundaries, but it is there in the sources, and it may be practical, as borrowing djifoa are large). We think that the stress shift in non-final borrowing djifoa may serve as a useful marker that something odd is going on when this happens.²⁹

Note that stress strongly constrains where a CV-V djifoa may appear if the VV is a doubled vowel.

Nothing in the description of complex predicates is new, though all language about borrowing djifoa comes from decisions taken in the 1990s after the 1989 edition of Loglan 1. Considerations about explicit stress markers

in an earlier footnote for determining whether an apparent cmapua unit before an initial pair of consonants is actually a cmapua unit enforces this already by causing the CV- in a CVC djifoa to fall off if it is initial and makes an initial pair with a following consonant.

²⁹We do not allow secondary stress on the penultimate syllable of a borrowing djifoa (the final syllable of the parent borrowing) or a pause after it if what follows the borrowing djifoa in the complex is not itself a legal complex. To quote an example where we actually ran afoul of this, it is not legal to articulate **aurmoykoo** as either **aurmo'yko'o** or **aurmo'y, ko'o**, because **ko'o** itself is not a legal complex. In **igllu'ymao**, the final syllable of the borrowing djifoa is stressed, but it must be because it is penultimate in the predicate word (ignoring the y syllable on which stress cannot fall); this is primary, not secondary, stress. **igllu'y, mao** is not permitted.

and syllable breaks are new but consistent with the logic of complex predicates as already defined: we do not want an illegal complex to become a legal borrowing by moving a syllable break so that it doesn't conform with a djifoa boundary.

2.4.7 Appendix: Alien Text and Quotations

The latest version of the parser does a complete pass with a checker for the phonetics, then does a pass with the lexicography and grammar checker. Thus, the phonetics checker needs to be able to recognize constructions with alien text (strong quotation with **lie**, foreign names with **lao**, foreign predicates with **sao**, onomatopoeia with **sue**, vocatives and inverse vocatives with **hue**, **hue**, and numbers with **lio**).

The parser recognizes the first four markers as inevitably followed by alien text. Alien text comes in two forms: it can begin with a double quote, end with a double quote, and contain any character but double quote, or it can consist of blocks of text containing any characters but whitespace, commas, and terminal punctuation and separated by the special word **y**. Blocks of alien text are set off with pauses initially and finally, if they are to be pronounced (and Loglan provides no advice on how to pronounce them). These pauses can be expressed by commas but do not have to be. When **hoi** and **hue** are followed directly by alien text (when addressing someone whose name is illegal in Loglan) the alien text must be enclosed in quotes: this is practical, as typos or grammatical errors might go unnoticed due to the parser accepting bad Loglan text as alien text.

The format for alien text is essentially that allowed to follow **lao** in 1990's Loglan (originally the constructor for Linnaean names: Steve Rice observed correctly that it was better thought of and used as a general foreign name constructor), though the **y**, to be inserted at whitespace was left unexpressed in writing in the original proposal. We require it to be expressed in the absence of quotes, and note that of course whitespace is to be read as **y**, in quoted alien text. If we want to refer to Einstein with his native spelling, we can style him **lao Einstein** and pronounce this as in German or English, where we must write **la Ainctain** if we want to import his name properly into Loglan. If we style him more elaborately as **lao Albert Einstein** we must remember (though this is optional in writing) that this is read **lao Albert y Einstein**. This proposal is not new for Linnaeans, in effect, but it is NEW as an implementation of strong quotation (the 1989 Loglan version is

quite unparseable in BNF or PEG formats) and allows NEW constructions in the cases of **sue** and **sao** (**sao** "ice cream", to be read **sao ice y cream** (also transcribable in the latter form).

The phonetic rules also support use of double quotes in **li...lu** quotations by recognizing **li** followed by a double quote followed by a phonetically well formed utterance followed by a double quote followed by **lu** (with possible intervening whitespace and comma pauses) as phonetically valid. Use of double quotes is optional (and the parser will keep track of correct nesting, which is not the case for the alien text quotes). The grammar will further enforce that what is enclosed in these quotes is a grammatical Loglan utterance. Similarly, the use of parentheses in the format **kie** (...) **kiu** in a parenthetical free modifier is supported, and can be nested.

2.5 There is a rationale, or "The original sin of Loglan?"

There is a rationale behind all of the ramifications of the Loglan phonetic proposal found above. I'm going to try to present the motivation behind all of this.

James Jennings has suggested that the attempt to make Loglan word forms recognizable by controlling patterns of C's and V's was a mistake, "the original sin of the language", as he put it. Whether this is the case or not, since our basic vocabulary is built on this principle, it is not something we can change without giving up and starting afresh. Personally, I think the Loglan phonetic system is actually rather charming, though admittedly baroque.

The key idea in the beginning was that there were to be three classes of words.

names: Names were to be recognizable because of their property of ending with a consonant, shared by no other word in the language, followed by a pause, so that the consonant could not be absorbed into a following word.

structure words: Structure words were to be recognizable because made up entirely of V, CV, and CVV units (Cvv-V being a later refinement). In any case, there would never be two successive consonants in a structure

2.5. *THERE IS A RATIONALE, OR "THE ORIGINAL SIN OF LOGLAN?"* 35

word, and structure words (and phrases made of structure words) would be for the most part flat evenly stressed sequences of these units. Vowel initial words presented special problems which required that one pause before them.

predicate words: The original specification of a predicate word was that it end with a vowel (so that it is not a name) and it contain at least one CC juncture (so that it is not a structure word; CyC being a later discovery). The beginning of a predicate word could then be recognized as CC- or CV(V)CC-; the end of a predicate word was to be recognized by penultimate stress. This immediately required the rule that a finally stressed structure word or structure word phrase preceding a predicate word must be followed by an explicit pause.

The phonetics of names were underspecified from the beginning, and the left boundaries of names presented difficulties, originally resolved by requiring that one had to pause at the beginning of a name as well, unless the name was preceded by a name marker, and forbidding phonetic copies of name markers from occurring in name words. Not allowing **la** to occur in a name is ugly.

The original predicate words had a simple flat structure: they consisted of chains of CCV and CVC units followed by a final CV.

The consonant clustering required rules about permitted initial pairs of consonants and permitted medial sequences of two or three consonants, which are preserved in our present scheme. All languages have such rules, and they are often rather arbitrary.

The formation of structure words and names required rules about vowel grouping. The double vowel rule (imposing stress) should be noted as having effects. Note that there was originally no vowel grouping in predicate words at all! We note that **ao** is a charming eccentricity which we should not try to change.

Notice that the simple scheme outlined so far allows easy syllable resolution for predicate words and structure words (and leaves names as a sort of phonetic black hole).

The next complication was the Great Morphological Revolution, with the attendant desire to build complex predicates out of units of the shapes CCV, CVC and also CVV – the last units presenting the problem that they do not enforce formation of a CC juncture. This caused the need for phonetic

hyphenization of djifoa (which were originally called “affixes”), with with **r** and **n** but also with **y** (the introduction of **y** doing some damage to the purity of our vowel system; I would like this vowel to have a more distinctive pronunciation, as John Cowan has already suggested). The construction of complexes is a complex enterprise, but the definition can be given and is unambiguous. There is weirdness attendant on the doubled vowel rule, which restricts the positions in which certain djifoa can appear, and also of course the fact that a CVC djifoa cannot appear in final position.

Even weirder and more wonderful was the attendant notion of borrowed predicate: any phonetic predicate word which cannot be understood as a complex! This occasioned the old **slinkui** test and the new CVC-**y** rule which replaces it, which prevent certain borrowings from being inadvertently read as complexes, and the arrangements for phonetic hyphens with **h** and syllabic consonants to “break” borrowed words so they cannot be read as complexes.

Later revisions or clarifications are

false name marker issues: the permission of phonetic copies of name markers in name words *if the name words are marked* and the elimination of common unmarked occurrences of name words (as for example unmarked vocatives). This was underway in the 90’s and completed by us after 2013. Our rules for recognizing left boundaries of names work, and they do allow **la** to occur in names.

the formal definition of the syllable: This was clearly needed in order to define what a correctly formed borrowed predicate was, and there was enough guidance in the Sources that my definition apparently essentially coincides with an already implicitly given definition. I further imposed the requirement that names should resolve into syllables.

doubled continuants: A minor point is that I decided to double all continuants used as syllabic consonants. JCB had suggested this.

orthography of names: Names with **la** are always to be written phonetically and cannot end with clusters of three consonants (this can usually be fixed because one of them is generally a continuant and can be doubled). Names are spelled differently than in the Sources rather often, but no names have been rejected other than ones with **q,w,x**, which were converted to foreign names. The interpretation of the old Linnaean name operator **lao** as handling foreign names made it possible

2.5. *THERE IS A RATIONALE, OR "THE ORIGINAL SIN OF LOGLAN?"* 37

to resolve the issue of whether names with **la** should be phonetic or preserve foreign orthography.

doubled vowels forbidden in borrowings: A minor point was that it proved convenient to forbid the doubled vowels with their attendant stress from occurring in borrowed predicates. If stress is not given explicitly, it has to be deduced from the orthography, and the doubled vowels made this harder. Only one word had to be revised. Doubled vowels in complexes are annoying but we have a fair number of them in the vocabulary, so we must view them as adding eccentric charm.

borrowing djifoa developed: We developed the full formal definition of djifoa derived from borrowings, which is plain weird. Everything we do with this follows from points in the sources. The actual shape was agreed on by the Academy in the 90s and is exemplified in the dictionary. The permission to pause after a stressed borrowing djifoa is described in L1, though on the basis of an older definition of borrowing djifoa: the stress shift in stressed borrowing djifoa is a logical consequence of prior rulings but may never have been noticed.

alien text constructions treated systematically: The incorporation of non-Loglan text and speech into Loglan in foreign names (described in the 90s), strong quotation (we changed this to follow the same pattern as foreign names), and foreign and onomatopoeic predicates is made systematic.

vowel grouping rationalized: Vowel grouping in structure words is unchanged. JCB's suggested rule for names makes no sense (right grouping of arbitrarily long vowel strings). I made my own algorithm, left grouping with a lookahead of three, preferring to group mandatory monosyllables and **ii**, **uu** together where possible (the latter to avoid occasions for the doubled vowel stress rule), and more weakly preferring to group the optional monosyllables together. This rule is applied in names and in borrowings. The ability to force a different grouping in names with a hyphen (not the original close comma) is preserved.

foreign letters eliminated: The letters **q,w,x**, introduced with strange pronunciations in the middle of the history of the language, had already been eliminated from all predicate words in the 90s and we completed their elimination from the language (except in alien text).

Finally, I am very pleased to be able to present a genuine phonetic parser for Loglan as a mode of the ordinary parser. To achieve this, I had to make sure that one could write any Loglan utterance in a form with explicit stresses and syllable breaks where desired (non-final stresses being treated as a species of syllable break and so following the syllable rather than the vowel) and with no whitespace except explicitly comma marked pauses. LIP did not actually support explicit commas in all cases where word breaks were mandatory! Because of the phonetic parser, we no longer need phonetic notation for the language (the one used in our Sources is a bit strange) and we have a native way to indicate stress which can be used in written text for emphasis. It is also charming that the rule about pausing after a finally stressed *cmapura* before a predicate word, which goes back to the beginning of the language, is actually expressible in the orthography!

Chapter 3

Lexicography

The subject of this chapter is the Loglan “word”. As we will see, this does not precisely accord with the definition given in NB3, though that is a place to start. In NB3, JCB said that the fundamental hallmark of a word is that one cannot pause in the middle of it (without changing its meaning). In Loglan 1 (1989), however, JCB allows pauses after borrowing djifoa in predicate words, which nonetheless clearly remain words.

It should also in general always be possible to pause before and after a word. Further, our intention (perhaps not perfectly realized) is that when whitespace represents a pause, not forced by phonetics to represent a pause, and omission of the whitespace would form a word, mere whitespace is not permitted: a comma-marked pause will be as a rule required in such a situation.

In earlier material of ours, we have maintained that classes PA and PANOPAUSES (in which we do allow pauses) are word classes. We now think the analysis suggests that the word units in these classes are smaller (though not always monosyllables) so we expect not to list these as exceptions to JCB’s criterion in this document. We still view pauses in class NI between digits as not being word breaks, and so constituting an exception. We introduced the possibility of pausing and resuming in acronyms, which we will present here as an exception; we also, however, now deprecate the use of Loglan 1989 style acronyms in favor of using names.

In the sister language Lojban, it is said that it is possible to pause anywhere in a stream of unit *cmapua* without affecting meaning, and so that there are no multisyllable *cmapua* words. We believe that this is an elegant situation, but also that TLI Loglan definitely has not achieved it. We do

have multisyllable *cmapua* words, and we will give an analysis of these in this document. There are some multisyllable words which *must* be multisyllable words on pain of ambiguities, and there are some which we view as multisyllable words as a matter of style: some unit *cmapua* seem to us clearly to be affixes in the proper linguistic sense rather than words, and we do not see value added in being allowed to pause before them.

3.1 Name Words (and Alien Text)

The description of name words is simple. These are the name words and acronym words of the phonetics section.

Djan is a word of this class. **DaiNaizA** is a word of this class. Words of these classes are followed by pauses.

The use of the 1989 (or earlier) acronym words is now in my mind deprecated (though still supported): I prefer the use of name words proper obtained by suffixing **-n** to the original acronym words in place of these (as for example **DaiNaizAn**), and I have made grammatical arrangements to support this in the case of dimensions. It should be noticed that independently of phonetics we did make the decision that acronyms should be names, not predicate words, which is **NEW**.

I regard the name marker words and alien text marker words as words, separate from the names or alien text which follow them. This analysis might be open to debate. A chunk of alien text is not by its nature a Loglan word! The **y** which separates chunks of alien text is I suppose a Loglan word.

I have been more conservative in this implementation of Loglan grammar (built on the phonetics proposal) about how many name markers there are.

- la:** The article which constructs names. A name marker. It can also appear as an article in descriptions.
- hoi:** The principal vocative constructor. A name marker. It can also be followed by other grammatical forms. The words of social lubrication are also vocative markers, but are not name markers.
- hue:** The inverse vocative constructure. A name marker. It can also be followed by other grammatical forms.
- ci:** A general “verbal hyphen”. Because it can be used between items in serial names, it is a name marker.

liu: The word quotation operator. A name marker. The non-Loglan word quotation marker **niu** is not a name marker, and is not likely to be used before name words.

gao: A proposed operator which converts words to letterals. Because it can be used with names, as in **gao Alef**, it is a name marker.

mue: A new name marker: this is used before a name word used as a dimension. **mue** appears as a unit in acronyms of the 1989 form, but in those contexts it is not a word.

Here are the alien text marker words and **y**.

lao: The foreign name constructor.

lie: The strong quotation operator. It should be noted that our grammar for strong quotation is totally different from the 1989 grammar of strong quotation: it is modelled instead on the 1989 grammar of **lao**.

sao: The foreign predicate constructor.

sue: The onomatopoeic predicate constructor: **sue miao** to meow. **sue sss** to hiss, **sue ccc** to shush.

hoi, hue: The vocative and inverse vocative markers can be used as alien text markers. When they are so used, the following alien text must be enclosed in quotes in writing. This avoids the problem of text with errors in it which should fail to parse parsing unintentionally as alien text.¹

y: A word used to separate blocks in a single item of alien text.

3.2 Predicate Words

The description of predicate words (in the semantic rather than phonetic sense) is straightforward, though it is not quite as simple as “the phonetic predicate words are the (semantic) predicate words”.

¹**lio**, the numerical description former, should also be an alien text marker, possibly with the same quotation rules as these. Added as an alien text marker 5/11/18 without required quotes.

To begin with, the phonetic predicate words (including those with internal pauses after stressed borrowing djifoa) are words. To give an account in which units ending with stressed borrowing djifoa were themselves words would be a complication, not a simplification. Such an account would be made easier by our requirement (not stated, or not made clear in the original proposal) that what follows such a pause must itself be a well-formed complex, but this sort of account would add no value to the grammar.

Numerical predicates are pause-free NI cmapua phrases (described below) followed by one of **ra**, **ri**, **ro**. N-**ra** is the predicate of N element sets. N-**ri** represents the predicate “X is the Nth item in series Y”. N-**ro** (a 1990’s innovation) is used to qualify other predicates: N-**ro pred**a means Nth most **pred**a (**nero gudbi** is “best”, **toro gudbi** is “second best”)². These are genuinely words because pausing in the middle would produce a description of a certain quantity of items described by a different predicate word. **tetora** is the predicate “is a set with thirty two elements”; **te tora** is an indefinite description, “three pairs”. In 1989 Loglan, numerical predicates had to be penultimately stressed; we have partially implemented this requirement in our parser (there is some freedom of stress placement in some cases). In any case this is not required to recognize where these words start or end. More information about the structure of these words will be found in the discussion of NI phrases later in this document.

The words

bia: (is part of),

bie: (is a member of (a set)),

cie: (is less than (math)),

cio: (is greater than (math)),

bi: (is defined as)

are all predicate words semantically, though they are structure words phonetically. They form a grammatical class BI of “identity predicates” (not a terribly accurate description).

I propose adding to the class BI all the forms obtained by prefixing **nu**, giving converse operators (my parser allows this). These converse forms are

²Some discussion of the place structure of these predicates is wanted.

words (the grammar does not allow **nu** to be applied to identity predicates as to normal predicates).

The words

he: (interrogative predicate; a sentence with a **he** in it is a question with a predicate answer),

dua: (first free predicate variable),

dui: (second free predicate variable),

buu: (first bound predicate variable),

bui: (second bound predicate variable)

are grammatically ordinary predicates, though phonetically structure words. None of them are really very ordinary predicates!

The acronyms, which were predicates in 1989 Loglan, are treated as names under our proposals (and in fact we suggest phasing them out and using names for their purposes).

3.3 Structure Word Classes

Most but not all cmapua words are single syllables or disyllables (one unit cmapua). Many compound forms often written without spaces, such as **lemi**, “my”, actually fall apart into two words. But some do not. Of these, some simply **cannot** be so viewed, and some we think do not make sense as compound words. We will discuss all these cases.

3.3.1 Tight logical connectives: CA roots

There is a series of logical connectives which must be presented first, as words (or affixes) of this class appear as components of elements of many other classes (including some complex logical connectives!)

The root words of class CA are

ca: and/or

ce: and

co: if and only if

cu: whether or not

nucu: converse to **cu**

ciha: interrogative quantifier

ze: mixture

CA roots may be prefixed with **no**, indicating negation of the first connected item, and/or suffixed with **noi**, indicating negation of the second connected item. Such a structure is called a CA core (a CA root optionally decorated with initial and/or following negations).

3.3.2 Letters, acronyms, and pronouns

A Loglan upper case consonant letter is **Cai**. A Loglan lower case consonant letter is **Cei**. A third series **Ceo** is provided for lower case Greek letters. Further series **Caiu** and **Ceiu** are provided: **QqWwXx** are **Kaiu**, **keiu**, **Vaiu**, **veiu**, **Haiu**, **heiu**. What the other new letters are, who knows?

A Loglan lower case vowel has the form **ziV**, and the upper case form is **ziVma**. The old style forms **Vfi** and **Vma** are fully supported in the parser, though we are not fond of them. These include **yma**, **yfi**. The **Vzi** form for lower case Greek letters is supported. The VCV letterals are multisyllable *cmapua* words.

The primary use of the letters in Loglan is *not* as names of phonemes but as **pronouns**. As a pronoun, a letter refers back to a recent argument with the same initial letter. There is a convention favoring using capital letters to refer back to proper names and lower case letters for general descriptions.

There is a further class of atomic pronoun words

tao: (that [of situations]),

tio: (this [of situations]),

tua: (???tu ze da. this may be obsolete),

mio: (we (first + third), independently),

miu: (we (first + third) mass),

muo: (we (first + second+third) independently),
muu: (we (first + second + third) mass),
toa: (this [of text]),
toi: (that [of text]),
too: (you, plural, independently),
tou: (you, plural, jointly),
tuo: (you and others independently (2+3)),
tuu: (you and others (2+3) mass),
suo: (self),
hu: (interrogative pronoun),
(ba, be, bo , bu): series of indefinite [quantified] pronouns,
(da, de, di do du): the series of old-style definite pronouns,
mi: (I),
tu: (you),
mu: (we (1+2) mass),
ti: (this),
ta: (that),
mo: (we (1+2) independently)

The anaphora convention for the series **da, de, di, do, du** can be read about in L1. The idea is that these words live on a stack in alphabetical order (those that are not already in use) and the *n*th description back in the text not already bound to a pronoun will be bound to the *n*th letter on this stack when needed. It seems rather baroque but very simple cases can surely be used correctly. We note also that the existence of the digit-suffixed forms should make it easier to use this system.

The general class of pronoun words consists of letters or other pronouns, optionally suffixed with **ci** followed by a NI0 unit (usually a digit; see the section below on numerals and quantifiers). It is advisable to pause after a digit suffixed pronoun like **dacine** (because **ci** is a name marker; with a little work in the parser I might be able to ensure that such occurrences of **ci** are not tagged as name markers). The numerically indexed pronouns are multisyllable cmapua words. It is very important to notice that for us a pronoun is a **single letter**, possibly suffixed with a single digit. Multiletter variables lead to horrible ambiguities which do serious grammatical damage. Multiletter pronouns are in fact supported by LIP but there is language in NB3 which suggests that JCB did not intend to have them, and we *strictly forbid multiletter pronouns (repetition deliberate)*.

The reason that it is vitally important **not** to allow multiletter pronouns is that the use of a sequence of individual letters as a sequence of pronoun arguments without the inconvenience of having to pause after each one is grammatically far more important than any use of sequences of letters as single pronouns or acronyms.

Further letter words, which may be used as pronouns, but to which we may not attach numerical suffixes, are generated by **gao** followed by a single well-formed word, either a name, a predicate word, or a consonant initial unit cmapua (CVV or CV). This is a proposal of John Cowan, intended to provide names for letters in alien alphabets.

3.3.3 Remarks on acronyms

An acronym is a sequence of letter names (possibly abbreviated in the case of vowels to zV – not to just V as in older versions of the language – which eliminates distinctions of case of course; corrections of V to zV in acronyms may be required in old texts), and number names (atomic quantifier words or numeral units), beginning either with the acronym marker **mue** [a proposed feature] or a letter (possibly abbreviated) and having more than one component (the dummy **mue** allows the formation of one letter acronyms and also of numeral initial acronyms without confusion with numerals or letterals). Acronyms are used to form dimensioned numbers (as discussed below) and to form acronymic names (no longer acronymic predicates – a proposal of course). The initial marker **mue** ensures that dimensioned number acronyms are not confused with sequences of pronouns, and the fact that acronymic names are **names** ensures that they are head marked in a way which ensures

that they cannot be confused with sequences of letter pronouns. Acronyms must always be marked with **ci** when used as components of serial names or name-final descriptions. A pause, terminal punctuation, or end of text is required after an acronym (so it can never attempt to consume a following letteral pronoun). One can pause inside an acronym and resume if the pause is immediately followed by **mue**; this corrects for problems of resolution of sequences of letterals, especially where the VCV forms are involved.

We add as a footnote a remark on why we do not like the VCV letterals. When VCV letterals are used in acronyms, as in **la daiafi**, the analysis of this into phonetic *cmapua* units has to be **daia-fi**, not coordinated with the semantic analysis into **dai-afi**. I did take the trouble to make sure that though one must pause before VCV letterals where they appear as words rather than acronym components, one does not need to explicitly comma pause; they are treated in the same way as vowel-initial predicate words.

We currently propose that the use of acronyms be replaced by the use of actual name words, formed by appending **-n** to the legacy acronyms, and we have made grammatical arrangements to support this usage.

3.3.4 Numerals and quantifiers

The numerals in Loglan are

ni: (0),

ne: (1),

to: (2),

te: (3),

fo: (4),

fe: (5),

so: (6),

se: (7),

vo: (8),

ve: (9).

Other words of the atomic quantifier word class NI0 are

- kua:** (division)
- gie:** (left bracket),
- giu:** (right bracket),
- hie:** (left parenthesis),
- hiu:** (right parenthesis),
- kue:** (inverse division),
- nea:** (unary minus sign) ,
- nio:** (subtraction),
- pea:** (unary plus sign),
- pio:** (addition),
- suu:** (root),
- sua:** (exponent),
- tia:** (times),
- zoo:** (double prime),
- zoa:** (prime),
- pi:** (decimal point),
- re:** (more than half of (quantifier)),
- ru:** (enough of (quantifier)),
- hi:** (close comma),
- ho:** (interrogative quantifier)

The closely related RA class contains

ra: (all),

ri: (few),

ro: (many);

these words are distinct because they have a different meaning when they appear as a suffix to a quantifier word (a quantifier word with a suffix with the phonetic shape of a RA word is a numerical predicate, for which see below).³

The SA class of quantifier prefixes consists of

sa: (about/approximately (prefix to a quantifier, by itself **sara**),

si: (at most, prefix to a quantifier, by itself **sine**),

su: (some/any/at least (quantifier prefix) by itself **sune**),

sinoi: (more than; a prefix to a quantifier, by itself **sinoine**???) new proposal),

sunoi: (less than; a prefix to a quantifier, by itself **sunoira**???) new proposal)

sanoi: supported by the grammar, and its meaning is deducible, but seems not likely to be used.

The SA-**noi** forms are multisyllable words or units in multisyllable words: all uses of **-noi** are as affixes.

We moved **ie** (who/what/which?) to class SA and eliminated all special references to it as a class. Note that it could attach to somewhat higher level argument classes in the old grammar, but it can still attach to them in the form **ie me** under the new arrangements. In fact, any word in class SA other than **ie** itself can be prefixed with **ie** to give a new element of class SA (this was needed to support **iesu**, which appears in Notebook 3). Further, **ie** may be succeeded by a pause in all cases; phonetics officially forbids a “word” in the proper sense which contains VV units and other sorts of unit cmapua.

³This dual use of the RA words has been corrected in Lojban, but we believe we are stuck with it: it is just one of the peculiar charms of the original Loglan. It seems possible to us that it might be wise to put **re** and **ru** in this class as well. Done 5/11/18.

We give semantics for these words briefly, but we do not envisage incorporating any official grammar of mathematical expressions into TLI Loglan; such a grammar might be desired by a group of users of the language, and they can develop their own for local use.

We handle the items **ma** and **moa** (00 and 000) differently than in earlier descriptions of the language. We define a class NI1 of numeral units consisting of a numeral (any word of class NI0 but this really makes sense only for the digits⁴ followed optionally by **ma** then optionally by **moa**, and a digit may optionally follow **moa**. D **ma** means D followed by two zeroes; D **moa** means D followed by three zeroes. D (**ma**) **moa** n means D followed by (2+) 3n zeroes. Originally, **ma** and **mo** were words of class NI0 meaning 00 and 000. **mo** is overused for other purposes, so we changed it to **moa**, and the use of an exponent seems better than repeating it. Replacing **mo** with **moa** is occasionally necessary in old texts.

A quantifier core (class NI2) is a sequence linked by CA cores of items of the following kinds (the items linked may further optionally be suffixed with **noi**):

SA: A SA word.

numeral block: A sequence of one or more NI1 words, with internal whitespace or explicit pauses permitted. It may optionally be preceded by a SA word.

RA: A RA1 word, which may optionally be prefixed by a SA word (this last option is a change from 1989 Loglan). A RA1 unit is a RA word suffixed with **mo** and/or **moa** optionally followed by a numeral, to give forms with meanings like “several hundred”. Question: how do we say “several dozen”? Or do we? It is important to note here that **sara**, for example, is not a numerical predicate, but a quantifier; the 1989 Loglan predicate **sara** becomes **sarara**. Replacements of things like **sara**, **sira** with (resp.) **sarara**, **sinera** is an occasional correction needed in old texts.

A general quantifier word has a quite complex definition. It begins with a quantifier core as described above. This may optionally be followed by

⁴You live and learn: in the Visit I found a need for forms like **rimoa**, a few thousand, so there is also a class RA1.

an acronym which must start with the marker **mue** [or by the word **mue** followed by a consonant final name word]; if this is present it is the last element in the word and is followed by end of text, terminal punctuation or an explicit pause. There is a final option of appending **cu**. Old Loglan texts will not have the marker **mue** before dimensions; this may need to be inserted.

General quantifier words are regarded as multisyllable *cmapua* words, even when they contain pauses between NI1 units.

The suffix **cu** (a late proposal of the last Keugru) generates indefinite mass or set descriptors from quantifiers (which are themselves grammatically a species of quantifier). I have to think carefully about whether this construction really describes a set as JCB says or a mass object; JCB, especially in later periods, tended to confuse the two.

The acronym suffixes create dimensioned numbers. The initial marker **mue** is a proposal of ours.

Quantifiers have important grammatical uses in the language, to be revealed below. This is quite a separate issue from having a complex internal grammar of quantifiers/numerals, which we avoid⁵. The word “mex” (abbreviating “mathematical expression”) is used in the grammar section for quantifier words.

3.3.5 Tense/location/relation operators

The root words of this class (which we call PA words (or phrases) for short) are

gia: (time free continuous tense, -ing),

gua: (timeless habitual tense),

pia: (past continuous tense, until [before terms]),

pua: (was habitually -ing, continuous past tense),

nia: (continuous present tense, during [before terms]),

nua: (am now habitually -ing, continuous present tense),

⁵2/3/2018 I am considering some simple grammar of quantifier cores. Not implemented, I am just thinking about it.

biu: (possibly, under conditions X [before terms]),

fea : ...happens in the same possible world(s) as...(actuality, in the sense of Kripke models of possible worlds). Not necessarily an official part of Loglan.

fia: (will be -ing future continuous tense, since X [before terms]),

fua: (will habitually be -ing, future continuous tense),

via: (throughout a place of medium size),

vii: (throughout a small place),

viu: (throughout a large place),

ciu: (X ga Y ciu Z means Z ga Y as much as X ga Y) [left here for the moment but actually moved to class KOU in 3/9 fix],

coi: (according to rule X),

dau: (probably, likely under conditions X),

dii: (for, on behalf of X),

duo: (by method X),

foi: (X foi Y, X must Y, X ga Y foi Z, X must Y under conditions Z – Y a predicate),

fui : (should, same structure as foi),

gau: (can (same structure as foi?)),

hea: (by, with the help of, X),

kau: (can, is able to (structure of foi)),

kii: (with/accompanied by X),

kui: ...is accessible from...(in the abstract sense of Kripke models, possible worlds). Not necessarily an accepted part of Loglan. I am now quite in favor of using this to build null prepositional phrases in order to force an indefinite variable such as **ba** to have a larger scope, as in **Mi djano lezo ba mormao la Djan, guo kui raba**, “I know who killed John”, which literally means “I know for each x to what extent x killed John”; without the **kui ba** it would just mean “I know the extent to which it is true that someone killed John”.

lia: (like, in the way that – I suggest that X ga Y lia Z means that X ga Z as Y ga Z , but X ga Y lia lepo Z ga W means X ga Y as Z ga W),

lui : (for, in order to please X),

mia: (subjective subjunctive, mia lepo X = were X the case),

mou: (more than, structure of *ciu*) [left here for the moment but actually moved to class **KOU** in 3/9 fix],

nui: (may/is permitted to, structure of *foi*),

peu: (as for/concerning X), *roi* (X *roi* Y = X intends to Y ; X ga Y *roi* Z = X intends to Y under conditions Z),

ru: ...obligates/makes it necessary that... from a counterfactual proposal. Not in the dictionary; not necessarily an accepted part of Loglan.

sea : (instead of X),

sio: (certainly, certain under conditions X [before terms]),

tie: (with/through/by means of instrument X),

va: (in the middle distance, near X),

vi: (here, at X),

vu: (far away, far from X),

na: (now, present tense, at the same time as X),

pa : (past tense, before X),

fa: (future tense, after X)

pau: (ago): added 11/14/2015 to support its use in A First Visit to Loglandia. I am not convinced that we need this cmapua.

There is also a related small class of KOU roots

kou: (because (cause) of X),

moi: (because/in order to (motive) of X),

rau: (because (reason) of X),

soa: (because(logical premise) of X)

ciu: (X ga Y ciu Z means Z ga Y as much as/to the same degree as X ga Y)

mou: (more than, structure of ciu)

which can be prefixed with **nu**, **no**, or **nuno** to give additional forms which we call KOU cores (a root is also a core). A KOU core is a multisyllable cmapua word. A KOU core is in effect a PA root.

It is important to notice that **nokou lepo X** does not deny X; in fact, it asserts X and says that the main event happened in spite of X. Forms like **nukou** are converses: they are versions of “therefore X”. Forms like **nunokou** are versions of “nevertheless X”; X happens, but not because of the main event, rather in spite of it.

The forms with initial **no** are obligatory words: pausing between the syllables of the word **nokou**, for example can radically change the meaning of a Loglan utterance. **Mi cluva la Meris, nokou Mai bilti** means “I love Mary, but not because she is beautiful”. **Mi cluva la Meris, no, kou lepo Mai bilti** has the same meaning as **No, mi cluva la Meris, kou lepo Mai bilti**, “It is not the case that I love Mary because she is beautiful”: it is possible that I am not saying that I love Mary at all. In the second sentence, use of an explicit comma to indicate this unexpected usage is required. **Mi cluva la Meris, no kou lepo Mai bilti** will not parse.

The words **ciu** and **mou** were moved into class KOU, to support formation of negative and/or converse forms of these words which are described in Paradigm K on our web site, though they never seem to have been implemented in LIP. The new “causal connectives” **mouki** and **ciuki** (and relatives) created by this move may have uses (I like them very much!).

We propose (with implementation) that PA roots other than KOU cores may be converted with initial **nu-** and/or negated with final **-noi**: these forms enter into all subsequent constructions as PA units (these may be called PA cores). These forms may further be prefixed with non-logically connected NI words (also producing PA cores). The new forms are multisyllable cmapua words (except that pauses are permitted between digits in the optional initial NI segment): this is a case where pauses would not be harmful but it does not seem to me that the **nu** or **noi** are functioning as freestanding words: they are affixes in the proper sense of that word. The conversion and negation forms for KOU roots remain as before (and KOU cores may be further prefixed with non-logically connected NI to obtain more KOU cores). Replacing **nokou** with **kounoi** is probably a good idea, but this would involve extensive changes in existing text. A PA or KOU core may be further decorated with a qualifier of class ZI (**za**, **zi**, **zu**), still obtaining a PA or KOU core (to see the effects of these qualifiers on tense and location operators, see the dictionary).

The class PA of phrases used as tenses or standalone modifiers consists of strings of PA or KOU cores (which may or may not be separated by pauses), such strings possibly linked with CA cores to further such strings.

The class PANOPAUSES used in modifiers with an attached argument consists of PA or KOU **words** (strings of PA/KOU cores not separated by pauses) possibly linked with CA cores to further such words.

Strings of PA/KOU cores not separated by pauses are viewed as multisyllable cmapua words.

When a PA class phrase is followed by a PANOPAUSES class phrase, an explicitly marked comma pause must intervene.

We think that the intent of **Mi smarue pa, vi le kruma** and **Mi smarue pavi le kruma** is different. The pause in the first sentence must be explicitly marked.

If it is not desired to draw the distinction between PA and PANOPAUSES, the grammatical solution would be to require that a PA phrase used as a modifier must be closed with **gu** when followed by another modifier. In this case, the form of PA and PANOPAUSES phrases would be that given for PA phrases above, and each PA/KOU core would be a (possibly multisyllable) cmapua word.

The semantics of complex PA words will require a considerable essay, to be inserted here in due course. In particular, a summary of the location and tense words and their interaction with **-zV** suffixes is needed, since these

have some ad hoc features. **pazu** a long time ago versus **panazu** in the past for a long time interval is an example I insert to remind myself.

These phrases can be used as prepositions (followed by an argument) or as tenses in the broadest sense (followed by a predicate): note the difference in phonetic form between these two uses, indicated above. The word **ga** is a content free tense word not usable as a preposition. **ga** has other uses as well. Details of this will be seen in the grammar.

Where a PA word occurs as a suffix to another word form (with attached explicit pause), it is generally illegal for it to be replaced by whitespace followed by a PA word in turn followed by an explicit pause: where a PA suffix is legal, it cannot be replaced by a following PA word without an explicit pause being indicated. **Da na clivi, o na brute** (an example in L1) does not actually parse correctly with LIP because of lexer problems with APA words; an unintended **ona** is read. It parses correctly as written under the current parser. **Da na clivi, o na, brute** fails to parse under the current parser, because the given pause pattern is in danger of creating an **ona**. **Da na clivi, o, na, brute** does parse as intended, and so does **Da na clivi, o, nabrute**.

The system of tense and location words

Here we will lay out the system of compound tense and location words, indicating difficulties and possibly some suggestions for improvement.

The basic series of tense words is **pa, na, fa**, which mark present, past, future tense when they mark a predicate; **pa X, na X, fa X** mean before X, at the same time as X, after X, respectively.

A second series of tense words **pia, nia, fia** express continuous tenses. **pia preda** means “was preda-ing”. **pia X** means “until X”. **fia preda** means “will be preda-ing”. **fia X** means “since X”. **pia preda** means “was preda-ing”. **pia X** means “until X”. **nia preda** means “is preda-ing”. **nia X** means “during X (throughout)”.

A third series of tense words **pua, nua, fua** express habitual tenses. Their meanings are similar to those of the previous series, but they refer to events which often or usually happen during an indicated period rather than events which happen continuously during an indicated period.

These words can be compounded. Here are the dictionary meanings of compound tenses.

papa: had (been)... ed, sign of the past perfect tense.

pana: was/were then... ing, sign of the past coincident tense.

pafa: was/were going to..., sign of past progressive tense, English inexact

napa: has/have (been).../a..., sign of the present perfect tense; already

nana: am/are/is now... ing, sign of the present coincident tense.

nafa: is/are going to..., sign of present progressive tense, English inexact.

fapa: will have... (been) ed, sign of the future perfect tense.

fana: shall/will be then... ing, sign of the future coincident tense.

fafa: will-be going to..., describes an action which takes place after the (future) time being recounted.

These words can be qualified with the suffixes **zV**. Here are the dictionary entries.

pazi: just... ed/was just (now a), a modified tense operator; just before..., before event terms.

nazi: at/coincident with..., an instant in time; at the time when, momentary event clauses.

fazi: will immediately (be a)..., modified tense operator; just after, before event terms.

paza: lately/newly/recently... ed, not too long ago, a modified tense operator; shortly before..., before event terms.

naza: during/in..., in some short interval, with terms.

faza: will soon (be)/be about to/just going to.; shortly after, with clauses.

pazu: long before, some event, before clauses.

nazu: during, in some long interval, with terms; while, during some long event.

fazu: will eventually (be a), a modified tense oper.; long after, some event, before terms.

The dictionary definitions are not fully systematic. Notice that **nia** and **nazu** express different meanings of “while, during”. I think in spite of some ambiguity about **nazV** forms, that the **zV** operators do something uniform, qualifying the distance of the event from the argument (or the present in the case of tenses). **nazu** doesn’t say that the event actually is far from the present, but since it says the event is in a long interval around the present it permits a long distance from the present.

Continuous examples are also listed

piazu: for all that time until now, adverb and before preds; long-before then and until, with clauses.

niaza: while/throughout the short time, clauses.

niazu: while/throughout the long time, clauses.

fiazu: since, for a long time after, with clauses.

The basic series of location operators is **vi**, **va**, **vu**, at/near/far from.

The second series of location operators is **vii**, **via**, **viu**, throughout a small/medium/large sized place.

Here are the compounds listed in the dictionary.

vivi: around, in the place where, before terms.

viva: out of where, a short way, with clauses.

vivu: out of, for a long way, before terms.

vavi: into where, from nearby, before clauses.

vava: past where, nearby, before clauses.

vavu: away from, from near to far, before terms.

vuvi: into where, from far away, before clauses.

vuva: toward the place where, before clauses.

vuvu: past where, at a distance, before clauses.

Modifications with **zV** affixes:

vizi: right here/at this spot, before preds; at the spot where, with point like events.

vazi: near this spot/the spot where, of point like events, before predicates.

vuzi: far from this spot, before predicates; far from where, spatially limited events.

viza: in this place/small region, before preds; where, before spatially limited events.

vaza: near this place, before predicates; near the place where, of limited events.

vuzaz: far from this place, before predicates; far from where, of medium sized events.

vizu: in this place/big region, before preds; where, before spatially extensive events.

vazu: near this region, of extensive events, before predicates; near the place where, of extensive events.

vuzuz: far from this region, before predicates; far from where, of extensive events.

The difficulty here is that there really isn't a system as such – at least, if there is, it is only implicitly given. It is possible to extrapolate from this, and it is also possible to compare with the sister language Lojban, in which an effort has been made to systematize these issues.

Another point is the status of the qualifiers **zV**. These are affixes, and one of these terminates a PA core, for us. In LIP, these affixes seemed to terminate PA words. Thus we allow **pazicevuzu** and LIP does not.

It is clear that a lot more words are formally possible, both for my grammar and for LIP.

3.3.6 Connectives

There are numerous parallel classes of logical and causal connective words in Loglan. Here we are only talking about binary logical connectives like English “and”; the word **no** for the unary negation connective is the sole inhabitant of a separate word class of its own.

Logical connectives for sentence components

The basic series of connective roots is **a**, **e**, **o**, **u**, **nuu**, **ha**. These are words by themselves, but certain affixes can be attached to them to build a large class of words. One can add the prefix **no** and/or the suffix **noi** to an A root to obtain an A core.

We describe the class A of basic logical connectives⁶. The root is taken from **a**, **e**, **o**, **u**, **nuu**, **ha** (possibly with prefixed **no** and/or affixed **noi**, i.e., an A core). A complete PA word (a tense in the broadest sense) with no internal pauses or spaces may follow as a suffix; finally, if and only if a PA component is present, **fi** or a full comma pause must close the word. An A word may not be followed without intervening space by a PA word (with no internal pauses) then whitespace: this is purely a technical device to detect unclosed APA words in legacy text. It is worth noting that in the NB3 corpus, JCB appeared to be following a rule of closing IKOU words with commas as one would expect here (though not APA words).

All A words are preceded by explicit comma-marked pauses. The phonetic reason for this exists only when the words are vowel-initial, but the rule is enforced for all words of this class.

It should be noted that our treatment of APA words is a new proposal. These words present considerable difficulties in LIP, and have been abandoned entirely in Lojban. We have preserved them so far because they are common in the NB3 corpus and in the Visit to Loglandia, and because the related IKOU words, which present much the same difficulties of termination, are clearly not dispensable without doing some violence to the corpus. I have tried a couple of different solutions: my aims here are to produce a solution which will allow parsing of legacy text with minimum violence (some pauses) and which will impose no unexpected obligations to pause on a speaker who always closes APA words and their relatives with **fi**.

a means “or” (the inclusive and/or). **e** means “and”. **o** means “if and only if”. **u** means “whether or not”. **nuu** is the converse of **u** in the obvious sense. **ha** is the interrogative quantifier; an utterance with **ha** in it is a question which calls for an A word as an answer. Compounds built with **ha** are not excluded by the grammar but certainly would be odd.

Prefixing **no** has the effect of negating the part of the logically connected utterance before the A word. Suffixing **noi** has the effect of negating the

⁶The class is actually called A1 in most places in the PEG grammar, and this may sneak into this text.

part of the logically connected utterance after the A word.

Suffixing a PA word has different semantics depending on whether or not the PA word is a KOU word. X, **efa** Y means X and then Y while X **erau** Y means X because Y, and careful analysis reveals that the first is **fa** X, Y while the second is X, **rau** Y. This is a slip, but we suggest following Lojban and keeping it this way. The alternative would be to have **epa** mean “and then”.

We now describe other series of connectives. The ACI and AGE connectives consist of an A connective, with any pause or **fi** after a PA word omitted, followed by **ci**, **ge** respectively. These connectives differ from A in precedence; their uses will be discussed in the grammar proper. They must be preceded by a pause, just as in the case of A connectives.

The CA connectives are another related class (already briefly introduced above). They are not preceded by pauses. The CA root forms are **ca**, **ce**, **co**, **cu**, **nucu**, **ciha**, **ze**. A CA root or a CA root with a prefix **no** and/or a suffix **noi** is a CA core. The semantics of **ca**, **ce**, **co**, **cu**, **nucu**, **ciha** are analogous to those of the A forms (and adding the **no** and/or **noi** has the same effect). **ze** builds composite objects or mixed predicates; its semantics are entirely different.

A CA connective word may take all the forms of an A connective with the A root component replaced by the corresponding CA component. A preceding pause is not required. The word **ze** has uses which a general CA word does not have (it can connect arguments). I am contemplating the formal possibility of **zenoi** and wondering if it might be useful.

The precise extent of the system of logical connective words here is not the same as that supported by LIP, but it is close. The scheme here allows more CA words than LIP does; we will see if they are useful.

Sentence connectives and new utterance markers

The connectives given so far connect arguments and predicates. We now consider connectives which connect sentences.

The word **i** (always preceded by a pause) begins a new utterance, but can often be treated as if it were a high level logical connective meaning roughly **e**. Further words of the same class I can be constructed by appending a PA word as a suffix, which must be closed with **fi** or a comma pause. The same issue exists for semantics of IPA words that is discussed above for APA words. All words of this class are preceded by a phonetically mandated comma-marked

pause.

A word of the class ICA consists of I followed by a CA connective word. This is a logical connective acting between sentences (but it can also connect utterances at a higher level). Because it is vowel-initial, it must be preceded by a comma marked pause.

An I or ICA word cannot be followed by whitespace then a PA word (an explicit pause is needed to separate a sentence initial PA word from the I or ICA word).

There are further forms ICI and IGE constructed from words of class I or ICA by appending **ci** or **ge** (after removing closures on component PA words).

The closure of logical and sentence connectives with **fi** is a new proposal here (I used **gu** earlier, but it creates conflicts, and I have experimented with different pause conventions).

Forethought logical and causal connectives

The root forethought logical connective forms are **ka, ke, ko, ku, nuku, kiha**, each possibly followed by **noi**. The root KOU words are **kou, moi, rau, soa** [under a proposal also **ciu, mou**] (optionally prefixed with **nu, no** or **nuno** to give forms which we call KOU cores (roots are cores too)), of which we will have more to say later. The forethought logical connective words of class KA are either one of these root words, or a KOU core, followed by **ki** then possibly **noi**. These forms appear before the first of the two items connected, with **ki** or **kinoi** appearing between the two items. Forethought connectives can connect almost any grammatical structure that can be linked by logical connectives. Note that forethought analogues of APA words are not provided; they did exist in LIP and could easily be restored if wanted.

The force of the causal connectives such as **kouki X ki Y** is (for example) X and Y (because of X). **nokouki X ki Y** is (for example) X and Y (not because of (in spite of) X). Note that the initial **no** is not negating X or Y, they are both asserted!

The new connectives **mouki** and **ciuki** have fairly clear meanings: **mouki X ki y**, “X more than Y”. **Mi cluva mouki la Meris, ki la Selis**, “I love Mary more than Sally”. **Mouki mi cluva tu, ki tu cluva mi**, “It is more the case that I love you than that you love me”.

How these words are *used* will be discussed below in the grammar.

Semantics of APA and IPA connectives

The reader will have noted already that there are grammatical and phonetic hazards associated with the APA and IPA connectives. There are also questions of interpretation.

La herba ga rodja, ikou mi ga cuirduo hei

The plant grew because I watered it.
transforms to

Kou lepo mi ga cuirduo la herba guo hei ga rodja

Because I watered the plant, it grew.

But

Mi cuirduo la herba, efa hei rodja

I watered the plant, then it grew
does not mean the same thing as

Fa le herba rodja guo, mi cuirduo hei

After the plant grew, i watered it
but exactly the reverse!

There was an attempt in the 1990's to reverse the definitions of these words for tenses, so that **epa** would mean "and then". Our view is that we should acknowledge that APA and AKOU (and IPA and IKOU) connectives are treated differently, and preserve the meanings of words in existing text.

Note that the first example uses an IPA connective and the second an APA connective. Using **ekou** in the first example would produce a sentence saying essentially the same thing.

More subtle considerations arise with APA connectives (or ICAPA connectives) with connectives other than **e**.

We argue that A, **icanukou** B must mean "Either A, or B, *because A is false*", or at least we propose to rule that this is what it means.

La Djan fa cluva la Meris, icanurau, mei fa kecri

John will love Mary, or she will be sad (for this reason, that he does not love her).

Similarly, A, **icakou**, B means "Either A, or not-A, because B".

This has the interesting effect that

Mi cuirduo le herba, inocanukou, hei rodja

If I water the plant, it will grow (because I watered it)

has the actual folk sense of the English if...then... sense, which includes a causal claim. This does not change the fact that the natural thing to say in Loglan is just

Mi cuirduo le herba, inoca hei rodja

The Loglander has a less causal and more logical view of the world.

Similar analyses for **okou**, **ukou**, **nuukou** and similar words can be supplied. The general idea is that in B, ICAKOU C, the logical force of the statement is that of the A connective used; when a given truth value of B allows both C and not-C or excludes both of them, no causal claim is made for that value of B; when a truth value of B permits only one truth value of C, then an additional implicit claim KOU lepo (not-)C guo, (not-)B is being made (negations being applied as required by the logic, and only of course applying in the case where the hypothesis and conclusion are true). In the case B **icanukou** C, we either have B (permitting both C and not-C, so no causal claim is made) or not-B and C (excluding not-C) so we make the additional causal claim **nukou** lepo C guo, not-B: C is true because not-B. The treatment of ICAPA is similar mod the different way in which PA claims are processed, indicated above. This may seem asymmetric, but the construction is intrinsically asymmetric, allowing only the second component to be qualified with a causal operator.

3.3.7 Articles

The basic articles (constructors of definite arguments) are

lea: article for sets: the set of all things with property ...

leu: The particular set I have in mind of things with property...

loe: The typical...

lee: The one or more things I mean which actually are...

laa: The unique object which actually is... (the logical definite description).

le: The default article. The object(s) understood from context which the hearer will be expected to think have property X...

lo: The mass article (describes composite objects made of all the objects designated).

la: The article for proper names.

These are now all the words of this class. The former construction of complex words of this class by following the root with an optional pronoun followed by an optional PA suffix has been superseded by a modification to the grammar class **descriptn**.

The name constructor **la** appears in the list above but appears in special constructions as well. The precise ways in which names are handled in this grammar involve new proposals.

There is a special class LEFORPO consisting of **le**, **lo**, and the quantifier cores (NI2) which may appear followed by PO in the formation of abstract descriptions. Notice that no new words are involved. It is worth noting that **lepo** and related forms are not single words, though they are often written without a space, and so can be written **le po** or even **le, po**.

Details of the use of these classes belong in the grammar below.

lau, **lua** and **lou**, **lou** are paired forms beginning and ending unordered and ordered lists, respectively.

3.3.8 Constructions involving alien text and related articles (see the appendix to the Phonetics Proposal for some modifications)

In this subsection we introduce the articles which handle quotations and imported foreign text, and we also give the full constructions of arguments (and predicates) of this kind. The strong quotation construction that we give is a completely new proposal.

Any well-formed Loglan utterance X can be quoted **li X lu**. X may be preceded and followed by explicit pauses (commas) if desired (this is not required). Under the Phonetics Proposal, we have not yet restored quotation of serial names (which are not utterances by themselves, though they are when marked) using **li/lu**, though we may do so. **li** is not a name marker word. I am contemplating allowing **li** to quote a descpred followed optionally by a name (this construction may now be the basis of a vocative or inverse vocative) but this seems less likely to be needed. Quotation marks may be inserted after **li** and before **lu** (and must match: if in one place, then also in the other).

A single Loglan word X may be quoted **liu X**. This is the only context in the grammar where the phonetic class of structure words plays any role. In LIP it plays no role even here, as LIP apparently only allows **liu** for

actual *cmapua*. Lojban I believe only allows unit *cmapua* to be quoted; we admit that there are compound words, so we allow them to be quoted. A **liu** construction must always end with an explicit pause (a new proposal, concurrent with the master Phonetics Proposal). **niu** may be used instead of **liu** to explicitly signal that a quoted word, though phonetically acceptable, is not a Loglan word. The Phonetics Proposal allows **liu** to quote marked names (as **liu la, Djan** or even **liu la, Djan Braon**) and alien text constructions (as **liu sao word**).

One may refer to a letter (rather than use it as a pronoun) using the form **lii X**.

The further forms discussed here operate on alien text. Alien text will be a block of text beginning with whitespace or an explicit pause and ending with whitespace, an explicit pause (comma), or before terminal punctuation or end of text, and containing no commas or terminal punctuation otherwise. It may contain other symbols or non-Loglan letters. Initial and final whitespace must be expressed phonetically as a pause.

The article **lao** followed by one or more blocks of alien text, with blocks being separated by **y** set off with spaces (which must be pronounced as explicit pauses) if there is more than one block, forms a foreign name. Wherever names are to be written by “look” rather than as they are to be read phonetically in Loglan, **lao** should probably be used. This construction was originally presented as a construction for the Linnaean names of biology; it is a valuable observation due to Steve Rice that it has a far more general usefulness. We abandon all other aspects of JCB’s discussion of Linnaean names as such: the details of scientific terminology are not part of the purview of the Loglan grammarian.

sao followed by alien text forms a predicate. This is a way to import a foreign word directly. **sue** followed by foreign text intended to transcribe or suggest a sound forms a predicate meaning “makes that sound”. **sue miao** is to meow.

Now we present our strong quotation proposal. The basic idea is that a series of blocks of alien text separated by whitespace is quoted by placing **lie** before the first block and **y** before each subsequent block. This is an entirely new proposal, though it turned out to be accidentally similar to the last proposal for the **lao** construction. The original strong quotation method is not PEG parsable (it is not even BNF parsable) and I think has other weaknesses. I have removed complexities of my original strong quotation proposal and made it parallel to **lao**.

The bit in Alice with the multifariously nested quotation marks must be translated into Loglan using this quotation style!

In the Phonetics Proposal, we have omitted the qualifiers **za** and **zi** for quotation of text versus speech.

We further note that the Phonetics Proposal allows alien text to be enclosed in double quotes, with whitespace allowed to be quoted (but pronounced , **y**, of course). The Phonetics Proposal *requires* that alien text following **hoi** or **hue** be quoted, to avert the possibility of non-name Loglan text with typos or grammatical errors being read as legal alien text inadvertently. The Phonetics Proposal allows multiple blocks of alien text to be used after **sao** or **sue**, with or without quotes, as in **sao “ice cream”**, pronounced (and also permitted to be written) as **sao ice y cream**, a predicate meaning (of course) “ice cream”.

3.3.9 Assorted grammatical particles, somewhat classified

Here is a list of terminators and boundary markers: **ci**, **cui**, **ga**, **ge**, **geu** (**cue**), **gi**, **go**, **gu**, **gui**, **guo**, **guu**, **gue**, and also the new **guoa**, **guoe**, **guoi**, **guoo**, **guou** (or alternatively **guoza**, **guozi**, **guozu**). There is a proposal of a new particle **gio**. Variants **guiza**, **guizi**, **guizu** are provided for the alternative parser.

New right closers **guea guua**, **giuo**, **meu** have been added.

The particles **je** and **jue** mark linked term sets (very tightly bound lists of arguments (and modifiers, in a new proposal)).

The JI words construct subordinate clauses from arguments, modifiers or predicates:

jie: (restrictive set membership),

jae: (nonrestrictive set membership),

pe: (general possessive),

ji: (which/that (is) (identifying),

ja: (which/that (is) nonidentifying

nuji: (new 1/10/2016) converse of **ji**: can be used to set values of pronouns.

La Djan, **nuji Daicine** sets reference of the pronoun **Daicine** to John.

The JIO words **jio**, **jao** construct subordinate clauses from sentences (resp. identifying, nonidentifying) Variants of the JI and JIO words suffixed with **za**, **zi**, or **zu** are provided, matched with alternative closers **guiza**, **guizi**, **guizu**. This allows efficient closure (with forethought) of nested subordinate clauses.

The case tags, including the positional ones are listed:

- beu:** (patients/parts),
- cau:** (quantities/amounts/values),
- dio:** (destinations/receivers),
- foa:** (wholes/sets/collectives),
- kao:** (actors/agents/doers),
- juj:** (lessers),
- neu:** (conditions/circumstances/fields),
- pou:** (products/purposes),
- goa:** (greaterers),
- sau:** (sources/reasons/causes),
- veu:** (effects/states/effects/deeds/means/routes),
- zua:** (first argument),
- zue:** (second argument),
- zui:** (third argument),
- zuo:** (fourth argument),
- zuu:** (fifth argument),
- lae:** (lae X = what is referred to by X),
- lue:** (lue X = something which refers to X)

The operators of indirect reference **lae** and **lue** are a different sort of creature, which originally had the same grammar as case tags, but now have somewhat different behavior. The latter two operators can be iterated (and so can case tags, probably indicating that more than one applies to the same argument).

My opinion of the optional case tag system is that I would never have installed it myself, and it represents an extra layer of work for dictionary maintenance, but it is potentially usable and represents a large amount of work by our predecessors, so my intention is to leave it in place (and try to be good about assigning tags when I define predicates) and maybe maybe some day actually learn the case tags! The whole scheme is quite optional for speakers, though pressure to learn them would be imposed on a hypothetical Loglan community if many speakers actually used them.

The particle **me** constructs predicates from arguments. I believe the addition of **mea** was a mistake, as **me**, properly understood, already served its exact function. I'll write an essay on this eventually. A new closer **meu** has been provided to close **me** predicates (**gu** will still work).

The particles **nu**, **fu**, **ju** interchange the 2nd, 3rd, 4th argument of a predicate respectively with the first. These are called conversion operators.

The particles **nuo**, **fu**, **juo** eliminate the 2nd, 3rd, 4th argument place of a predicate respectively, stipulating that it is occupied by the same object that occupies the first argument place (these are reflexives).

More conversion and reflexive words are formed by suffixing a quantifier. The only meaningful ones as far as I can see would be numerals larger than 4 and **ra**, which would choose the last argument place.

Yet more words of this class can be formed by concatenating conversion operators and reflexives; they simply compose, allowing complex reordering and identification of arguments.

Words which form abstraction predicates are the short-scope **poi**, **pui**, **zoi** and the long-scope **po**, **pu**, **zo**. In each set, the words form predicates for events, properties, and quantities respectively. Additional words **poia**, **poie**, **poii**, **poio**, **poiu**, **puia**, **puie**, **puii**, **puio**, **puiu**, **zoia**, **zoie**, **zoi**, **zoio**, **zoiu** are also long scope abstraction operators but with different closure words, **guoa**, **guoe**, **guoi**, **guoo**, **guou**, the final vowel indicating which closure word is to be used. There is an alternative version of this proposal adding abstraction words **poza**, **pozi**, **pozu**, **puza**, **puzi**, **puzu**, **zoza**, **zoz**, **zozu**, with closure words **guoza**, **guozi**, **guozu**; it is thought that **poia** in particular might be confused with **po ia** (though I disagree, insisting that a considerable pause is required in **po ia**) and certainly three additional

sets are sufficient.

The uses of all these words will be revealed by the grammar.

3.3.10 Words which form free modifiers

The register markers indicate attitude toward the person addressed:

die: (dear),

fie: (comrade/brother/sister),

kae: (gentle as in gentle reader to an equal at a certain distance),

nue: (Mr Ms Mrs neutral and at a distance),

rie: (Sir, Madam, Sire, Honorable – to a superior)

They can be negated: there is no reason that we cannot address people nastily in a logical language.

The vocative marker is **hoi**. The inverse vocative marker (indicating the speaker or author) is **hue**.

The “right scare quote” is **jo**, which may be prefixed with a numeral. It indicates that previous text is not to be taken quite literally; the numeral would indicate how many words are in the scope of the **jo**. I notice that if a scare quote were to be applied to a quantity, it would have to be **nejo**. *soi crano*.

The paired words **kie** and **kiu** serve as spoken parentheses: include a well-formed Loglan utterance between them to form a free modifier. Actual parentheses can now be inserted after **kie** and before **kiu**.

Smilies can be spoken in Loglan: **soi X**, where X is a predicate, forms a free modifier inviting the auditor to imagine the speaker doing X. **soi crano** is literally :-) Loglan smilies are almost as old as the historical origin of smilies, I believe.

The freestanding attitudinal words of the original VV flavor, generally expressing emotions or attitudes, are

ua: (there! thats it! done! satisfaction),

ue: (indeed! oh! surprise),

- ui** : (fine! good! (pleasure)),
- uo**: (come now! look here! (annoyance)),
- uu**: (Alas! Sorry! sadness/sympathy/regret/not apology, that is sie),
- oa**: (moral obligation – it must be),
- oe**: (preferably),
- oi** : (permissibly, you may),
- oo**: (disapproving hmmm)[to be added!],
- ou**: (no matter (ethical indifference)),
- ia** : (yes), agreement),
- ii** : (maybe (tentative belief)),
- io**: (I expect that, apparently, moderate belief),
- iu**: (I have no idea!, ignorance, lack of belief or knowledge),
- ea**: (let's, I suggest...),
- ee**: (caution! careful! take care! [to be added]),
- ei**: (is it true that? forms yes/no questions),
- eo**: (please? will you? asks permission),
- eu**: (let us suppose that...(subjunctive)),
- aa**: (I see (what you mean)),
- ae**: (yes, I wish to (hope or weak intention)),
- ai** : (I intend to...Definitely...(strong intention)),
- ao**: (Yes, I want to, Ill try to...(moderate intention)),
- au**: (I dont care...indifference, absence of intention)

ie is not really an attitudinal, but an interrogative meaning “which”. (the words **aa**, **ee**, **oo** are not in the trial.85 list of UI words, though likely the parser handles them fine in LIP; I have added them).

Additional words with the same grammar are

- bea**: (for example),
- buo**: (however, on the contrary, but),
- cea**: (in other words, namely),
- cia** : (similarly), **coa** (in short, briefly),
- dou**: (given, by hypothesis),
- fae** : (and vice versa),
- fao** : (finally, in conclusion),
- feu** : (in fact, actually),
- gea**: (again, I repeat),
- kuo**: (usually, customarily),
- kuu**: (generally),
- rea** : (clearly, obviously, of course),
- nao**: (now, next, new paragraph),
- nie** : (in detail, looking closely),
- pae**: (etc., and so forth) ,
- piu** : (in particular),
- saa**: (roughly, simplifying),
- sui** : (also, as well, furthermore),
- taa** : (in turn, sequence),
- toe** : (respectively),

voi : (skipping details),

zou: (by the way, incidentally),

ceu: (anyhow),

sii : (evidently)

These words are discourse operators, comments on the way we are speaking.

The word **cao** emphasizes the next word. The grammar will not show this, as it associates attitudinals with the previous word or construction! Notice that one can use the phonetic stress markers to indicate stress in writing.⁷

The word **seu** (a proposal) has a semantic effect, though it is grammatically an attitudinal. It marks an *answer*. This is **significantly useful**⁸ for indicating that a predicate word given as an answer to a question is not intended as an imperative or observative; it may have other uses.

Finally, we have words of social lubrication,

loi: (hello),

loa: (goodbye),

sia: (thank you),

siu: (you're welcome, dont mention it),

sie: (sorry (apology))

The word **sie** (to be distinguished from **uu**, sorry in the sense of regret but not apology) is new. Cyril and I believe it reasonable that **siu** be a polite answer to **sie** as well as **sia**.

⁷The word **kia** is listed as having the effect of cancelling the previous word. I do not at the moment intend to implement this: a grammatical implementation would involve recognizing certain **kia**-final constructions as freemods, and there would be decisions to make about what the units cancelled were to be (it appears to me for example that entire quoted constructions would be cancelled, and **liu kia** would be a quoted word, but there would be other restrictions, basically to do with the fact that a cancelled unit could occur only where a freemod could be expected).

⁸serving to compensate for the fact that Loglan, unlike Lojban, does not have an explicit marker for the imperative; we further compensate for this by insisting that tense-marked gasents are observatives, not imperatives.

There is a proposal, currently implemented, that these words also be vocative markers but not name markers, so that one can say **Loa Djan** as well as **Loa hoi Djan**.

The attitudinal, discourse and social words (class UI) can be negated by preceding them with **no** or following them with **noi** (the use of **noi** is a tiny proposal).⁹

In addition, there are discursive operators firstly, secondly, lastly formed by suffixing quantity words with **fi**.

3.3.11 Negation

The word **no** is the logical negation operator. Initial **no** in attitudinal forms, KOU words, and subordinate clauses (as well as occurrences internal to some compound structure words) must be excluded from this grammatical class. Pauses after **no** may be semantically significant, because they cause word breaks, and also because of the possible use of **no** to negate an entire utterance rather than its first argument (which usually does not affect meaning, though it affects the parse of a sentence).

3.4 Essays on word-making, and on what a word is exactly

3.4.1 Borrowing predicates

The responsibilities of a Loglan user borrowing a predicate from another language for use in Loglan are outlined.

One first roughly transcribes the word into Loglan phonetics. One replaces foreign sounds with Loglan sounds. It needs to be free of bad consonant combinations which Loglan doesn't support; this could be fixed by inserting vowels or sometimes by doubling continuants. Doubled non-continuants need to be undoubled.

It needs to have a left boundary of the right form. If it begins with a permissible initial consonant cluster, this is handled. Otherwise, we need

⁹The ability to write “words” like **noia** (explicitly articulated as **no-ia**, and without a pause before the vowel initial **ia**) requires explicit overrides of the usual phonetic rules; I doubt that **liu noia** will parse, but this can be pronounced without pause.

to look after its initial (C)Vⁿ and see if a consonant cluster can be introduced. Appending **h** after a second single consonant as in **athomi** has been a frequent maneuver.

It needs to have a right boundary of the right form, which really amounts to being vowel-final: a vowel is added if necessary.

It needs to not be a complex. A vowel initial borrowing is of course never a complex. Doubling a continuant as in **hidrroterapi** can prevent a borrowed predicate from being a complex (and in this case also prevented the initial **hi** from falling off as it otherwise would, **dr** being an initial pair of consonants: this kind of gluing is another reason to introduce a syllabic consonant in a borrowing). Ensuring the presence of a sequence of three vowels would do this cheaply. A final sequence of three vowels will always work to prevent resolution into a complex, if the resulting stress is bearable to the hearer.

The non-Loglan speaker may need to adapt to the stress being in an unexpected place. Part of the art of the borrower into Loglan is to try to make the word sound reasonably like the original while meeting the requirements for a Loglan borrowing.

It is also permissible for a borrowed word to take one of the shapes of five letter Loglan primitive predicates, CCVCV or CVCCV; it is not permissible for it to resolve into multiple djifoa. We do require that there are no Loglan predicates of the primitive shapes which differ only in their final vowel, unless they are actually variations of the same word, as in the animal or cultural “declensions”. This is vital because the identity of the final vowel is suppressed in forming the five letter djifoa. Such a borrowing becomes in effect a primitive and can form djifoa like any other primitive.

It is worth noting strategies used in salvaging VCCV initial borrowings: we have used doubling continuants, and also used initial **h**.

There are semantic requirements to making a predicate of either sort: one has to decide on an argument structure and, if one is really kind, decide on assignments of case tags to the arguments.

3.4.2 Making complex predicates

The responsibilities of the Loglan user in making complex predicates are outlined.

No new five letter “composite” atomic predicates are expected to be made: the esoteric process by which they were made does not need to be discussed.

One might in theory make a five letter predicate as a borrowing as noted in the previous section. This should not happen often.

The maker of a complex should have a metaphor in mind. The components of the metaphor are then arranged in a suitable order (there might be some freedom in the order as well). One then chooses the right djifoa associated with the components. A borrowing has only one djifoa form, of course. Every primitive predicate (and any five-letter borrowing) has its five letter final form and its five letter medial forms with final **y**. Most of the primitive predicates have one or more three letter forms available, and the Loglan learner (and certainly the imaginary Loglan native speaker) should *know* the djifoa as part of the root vocabulary.

There are then certain restraints on the use of the three letter forms. One has to make sure that there is a CC junction. In fact, the only situation where there is a CC junction problem is if the first djifoa one has chosen is CVV, and followed by a CVx or CVCCx, and the problem is fixed by hyphenating the first djifoa. An **r** or **n** hyphen is used by preference. A CVV with a **y** hyphen should be used only before a borrowing djifoa (where this is mandatory) or if the intention is that the CVV djifoa represent the cmapua of the same shape. The presence of a borrowing djifoa of course ensures the presence of a CC junction. We note with horror the possibility of complexes beginning **CVVy(C)VⁿCC**, which can happen if a CVV djifoa is followed by a borrowing djifoa. **CVCy(C)VⁿCC** is not much more appetizing.

This is a good moment to note that some **CVr** and (under a proposal of mine) all **CVh** djifoa are reserved to represent CV cmapua. The legacy vowel letterals may not be used as djifoa, but the new ones are eligible: **ziaytrena**, “A-train”.

A CVC djifoa in initial position will have to be followed by a **y** hyphen if an initial pair of consonants would otherwise be formed (or if it is followed by a borrowing djifoa). A **CyC** sequence does count as a CC pair, as in **mekykiu**.

The CVV djifoa with repeated vowels that force a stress cannot occur except in final position or in penultimate position, followed by a monosyllable.

Where a CVV which is an optional monosyllable ends a complex, it may be the case that two possible patterns of syllabification and stress are possible for the complex.

The remaining obligations are aesthetic: make a reasonably short, pronounceable and even pretty word. Aesthetics may vary: this writer *likes* the word **likcke**.

3.4. ESSAYS ON WORD-MAKING, AND ON WHAT A WORD IS EXACTLY⁷⁷

There are semantic requirements to making a predicate of either sort: one has to decide on an argument structure and, if one is really kind, decide on assignments of case tags to the arguments.

3.4.3 Name words

The name words consist of the name words in the phonetic sense of the first section and the acronyms. One is required to pause after an acronym used as a name, and one is permitted to omit the explicit comma in writing under exactly the same conditions as after an ordinary name word. It is worth noting that a pause is also required after an acronym when it is used as a dimension in a quantity.

Contrary to statements in L1, we maintain that a Loglan name word should always be written as it is to be pronounced. Names written to look visually like their forms in other languages should be treated as alien text and turned into grammatical proper names with **lao**. Thus, **la Ainctain** is the native version of Einstein's name, but we can of course also write **lao Einstein**. The first must usually be followed by an explicit pause, while the latter may be followed by an innocent space – which will also be a pause, as stated in the rules for alien text. **la Einstein** is a legal Loglan name, but would be pronounced quite oddly.

Creating Loglan proper names is generally a process of transcription of a name from some other language. Transcribed names must resolve into Loglan syllables. One should notice that we do not allow double consonants except for syllabic consonants, and that syllabic consonants *must* be doubled. Further, a name may not contain more than two successive non-syllabic consonants at the end, though this may be fixed by doubling a continuant, as in **la Marks**.

It is conventional but not required to convert a vowel-final name from foreign sources to a Loglan name by adding **s**; there is nothing wrong per se with using another consonant, particularly if there is an etymological reason to do so. Loglan names can be made from predicates by omitting final vowels or (conventionally) by adding **n**. Another idea which I have encountered recently is to make an illegal complex ending with a CVC djifoa and use this as a name, which strikes me as a lovely idea.

3.4.4 Essay: what is a word?

Cyril Slobin asks me, what is a Loglan word? How does the hearer resolve a stream of Loglan sounds or letters into words?

JCB's answer in NB3 was that a word is a sequence of phonemes in the midst of which one cannot pause.

This is not perfect, but it is a good approximation. JCB himself defined an exception: one can pause in the middle of a predicate word after a borrowing affix! Cyril himself proposed an exception for long NI words (numerals): pauses, even comma marked ones, between NI1 units do not affect semantics.

Name words are reasonably easy to recognize phonetically (pause free sequences of phonemes, usually marked by an initial name marker word, ending unmistakably with a pause after a consonant). They certainly meet JCB's criterion; pausing in the middle of a name breaks it. We also view the name marker word as a word.

Predicate words are fairly easily recognized phonetically, starting with a characteristic $(CV^n)CC$ phonetic configuration and ending with a penultimate stress. They do not break into separate breathgroups except for JCB's exception of allowing pauses after borrowing affixes. Of course, one might heretically view a predicate with a borrowing affix as a kind of phrase, but I think it is really still a word. Similar remarks apply to John Cowan's **zao** construction, another way to build a complex predicate which actually allows internal pauses¹⁰.

More headaches about what a word is arise with *cmapua*. The Lojbanists have apparently arranged things so that one can pause anywhere in a stream of *cmapua* syllables without affecting meaning, so that the unit "words" are just unit *cmapua*. This is not true in TLI Loglan. JCB certainly thought that compound *cmapua* words existed in the language. I regard the members of certain large *cmapua* classes as words, and in most cases I have enforced the rule that one cannot pause inside them.

I make a list of classes that are inhabited by multisyllable *cmapua* words.

TAIO: this class includes multisyllable names of letters that do not fall apart.

A: This class includes quite complex logical connectives. One cannot pause

¹⁰The **zao** construction is not currently described in this document (it vanished when I installed the Phonetics section in the old Report document), but it is my intention to restore it to the grammar as soon as I have reinstalled it in the parser.

3.4. ESSAYS ON WORD-MAKING, AND ON WHAT A WORD IS EXACTLY⁷⁹

inside such a word. **noapacenoina** is a long example.¹¹

ACI, AGE, CA: relatives of A, similarly large classes of words in which breaks are not permitted (except as in the footnote under the previous entry).

I, ICA, ICI, IGE: again phonetically and to some extent semantically similar to A.

KA, KI: These classes include compound words, all fairly short, since we exclude PA-suffixing of such words.

NI: This is a large class of quantifier words, and I really do think that they are words, except that I allow pauses between NI1 numeral units. This does not mean that one can freely pause anywhere in a NI word; at many junctures one cannot, and certain constructions unequivocally close such a word. The related class of numerical predicates does not allow internal pauses.

Acronym: Acronyms are words (or in the case of dimensions, parts of NI words). One cannot pause in the middle of an acronym, and its boundaries are clearly marked (by **mue** or a name marker on the left and a pause on the right).¹²

DA: Suffixed pronouns are multisyllable *cmapua*.

PA: Pause free strings of PA cores are words. Series of PA cores linked with CA cores are now viewed as phrases.

LE: Compound articles such as **lemi**, **levi** were words under LIP (LIP allowed spaces in them but not commas) and under previous versions of my parser, but I have (at least experimentally) modified class **descriptn** so that things like **le mi hasfa**, **le va hasfa**, **le mi na hasfa** are actually read word by word. The sentence **le mi hasfa** is now an instance of the same grammatical construction as **le la Djan, hasfa**, which was not true in *trial.85*, though every learner may have thought so.

¹¹One now **can** pause inside such a word, next to a CA0 connective, but it is still clearly a word.

¹²added the ability to insert **,mue** into an acronym, so yes, one can pause, but it still looks like a word class.

JI: I allow **nuji**.

NU: Suffixed conversion operators such as **nufe**.

UI: NI F i discursives are words. Negative attitudinals such as **noia** might be viewed as words: **liu noia** (with following pause) should parse under the Phonetics Proposal.

BI: I allow forms like **nubi**, which are treated as words (**La Djan, nubi da** is parseable, but **La Djan, nu bi da** is not: **nubi** is semantically but not grammatically parallel to **nu blanu**.)

Other cmapua classes define words inhabited by one-unit cmapua (not necessarily one syllable, as some unit cmapua are disyllables).

This is actually not a terribly long list. Familiarity with the phonetics of names and predicates (admittedly quite nasty in its finer technical details, but usually quite manageable in normal situations) and the grammar of a few word classes will allow you to recognize the Loglan word.

It is important to notice, though, that while the recognition of name and predicate words is a matter of phonetics, recognition of the cmapua words is a matter of understanding the grammar. They do have a common phonetic property (most of them), in not admitting internal pauses, but they are not resolved using phonetic criteria.

Chapter 4

Grammar

4.1 Introduction

In this chapter, I describe the grammar proper (phonetics and lexicography being handled in the previous chapters) which is implemented in my provisional Loglan parser(s). There are more changes in the grammar than there are in the lexicography or the phonetics, but in all cases the intention was to make the language work, not to make essential changes.

A recent alternative proposal affecting closures of complex constructions is described here *as an alternative*. It is less conservative than most of my changes and probably requires separate attention.

4.2 Sentences, Free Modifiers, and Utterances

We will begin with the Loglan sentence, then work our way up to more general Loglan utterances (and the ubiquitous free modifiers), and down to the components of sentences, which are predicates, terms (arguments and modifiers) and various flavors of lists of terms.

In our discussion of the sentence, we will use simple examples of the sentence components mentioned above which are fully explained in later sections.

4.2.1 The basic SVO statement

The most basic Loglan sentence form consists of a subject followed by a predicate: we call this a basic SVO statement. Of course, a lot of complexity is hidden in these words, and there are some additional optional components.

The subject is a term list. Term lists, as we will see below, consist of arguments (noun phrases such as **da** or **le mrenu**) and modifiers (prepositional phrases) such as **vi le hasfa** or **na la Ven**). A subject is a term list which contains at least one argument, and no more than one argument which does not have a case tag (semantic or numerical).

The predicate may include one or more termsets (an object or objects) as a final component.

The differences between term lists and termsets will be discussed later. A termset may have more grammatical structure than a term list, which is just a concatenation of arguments and modifiers. But any term list is a termset.

The order of the components in this form of the Loglan sentence is thus SVO (where there may be a number of objects). It is perhaps a defect in our grammar (inherited from the earliest design of the language) that we parse this as [S][VO], lumping the objects in with the verb phrase, while the logical sympathies of the objects are actually with the subject.

Here is a simple example:

La Djan, cluva la Meris

And another:

La Djan, na la Ven, donsu le bakso la Meris

Here we have more than one term before the “verb” and two objects after it.

It is important for reliable parsing of the language that we enforce the rule that there is no more than one untagged argument in the subject. JCB expressed the intention in NB3 that there should be just one argument in the subject part of the sentence, while finding it easier to make it a general term list in the formal grammar. On the other hand, he exploited this feature of the formal grammar later in order to support SOV word order. We find that allowing more than one argument before the predicate is dangerous: an improperly closed previous utterance may grab arguments from the subject. To support SOV(O) word order, we add an optional component between the subject and the predicate: this is the new particle **gio** followed by a loose term list.

We give an example

La Djan, gio le bakso ga donsu le Meris

The same effect is achieved by

La Djan, zue le bakso ga donsu la Meris

in which we get permission for the box to appear before the verb by applying a numerical case tag.

A “sentence” like

***La Djan, le bakso ga donsu la Meris**

is permitted in 1989 Loglan but forbidden by our grammar.

In an alternative version of the parser, there is a further optional component of this class of sentence (the basic SVO statement) which is a new particle **gaa** (the “large subject marker” which may appear immediately after the subject and before any arguments with **gio**, and whose uses will be discussed later.¹

4.2.2 Subject-final statements and observatives

We now describe the other forms of the basic Loglan sentence, in which the subject is final or absent. These sentences are traditionally called “gasents” in Loglan grammar, and we will use this terminology.

There are two basic structures for a gasent. The general idea is that a gasent consists of a tense word (class PA) or the particle **ga** followed by an untensed predicate (class **predicate** with an initial **barepred**), followed by the particle **ga**, followed by a term list.

The qualification which divides the class of gasents into two subclasses is that what appears after the last **ga** must be either a subject or *all of the arguments (in fact, all terms after the verb) in the sentence*.

The first class of gasents consists of a tense word, followed by an untensed predicate (which may include final termsets), followed (optionally) by a subject prefixed with **ga**.²

The second class of gasents consists of a tense word, followed by an untensed “verb phrase” (no objects), followed by **ga** followed by a term list

¹In this and later mentioned occurrences of **gaa**, it seems it may be desirable to allow it to be preceded by an optional negation **no**, or more than one). Done 5/11/18.

²Until a parser upgrade implemented while I was writing this, the predicate in a **gasent1** had to be a bare predicate, ruling out, for example, logically connected predicates with A connectives. I corrected this to avoid problems with the distinction between imperatives and observatives.

(which may optionally be of the form (subject + **gio** + term list), but use of **gio** is not required).

There are two important points of difference from 1989 Loglan here.

In the first class of gasent, the structure (**ga** + subject) is optional. A sentence such as

Na crina

is read as an observative, “It is raining”, with subject elided: it is interpreted as **Na crina (ga ba)**. In 1989 Loglan, such a sentence would be an imperative. We require that imperatives be untensed. We regard the observative form as useful.

The further, and possibly more important change, is that we require that either exactly one (untagged) argument follow the **ga** in these V(O)[**ga**]S sentences, or else all terms after the verb (and so all untagged arguments) follow the **ga**. The reason we feel this to be important is that the place structure of a gasent could otherwise be changed radically at the very end of the utterance by supplying two initial arguments rather than the expected one.

Under the alternative parser, gasents may optionally be adorned with an initial occurrence of the large subject marker **gaa**.

4.2.3 Sentences

The Loglan unit sentence is one of the following:

statements: Either an SOV statement or a gasent is a unit sentence.

imperatives: An untensed predicate is an imperative sentence. A term list consisting only of modifiers followed by an untensed predicate is an imperative sentence. These are species of unit sentence.

Under the alternative parser, imperatives may be marked with the large subject marker **gaa** just before the untensed predicate.

It seems obvious that **Na la Ven, prano** would be an imperative in 1989 Loglan, but this was not discussed anywhere. The formlessness of the initial term list in the SVO statement as described in the original Loglan grammar made it hard to see this as an important special case.

forethought connected sentence: A forethought connected sentence consists of a KA core followed by a Loglan sentence, followed by **ki** or **kinoi**

followed by a unit sentence or a sentence with head terms. This kind of sentence may optionally be multiply negated with initial occurrences of **no**. This is a species of unit sentence.

This is revised from a rule carried forward from 1989 Loglan until very recent versions of my prover which allowed the final component of a forethought connected sentence to belong to the very general class of utterances **uttA1**, which contains many sorts of sentence fragments normally uttered as answers to questions, as well as the two kinds of sentences indicated. To our minds, this situation was bizarre. In parsing the Visit to Loglandia we found a number of grammatical errors not detected because of the use of **uttA1**.

The unit sentence has an additional optional component: it may be prefixed with one or more “negheads”, either an occurrence of **no** followed by **gu**, or an occurrence of **no** followed by a pause, where the **no** cannot be absorbed into a adverbial modifier (**no** of class NO2). This is the last shadow of the “pause/GU” equivalence found in 1989 Loglan: here a pause may change the grammatical structure of a sentence (usually by causing **no** to negate the entire sentence rather than the first argument in the sentence) but not the semantics (the logical effect of the two analyses of the sentence is the same). Allowing negheads to attach to unit sentences as well as to class **uttA1** is part of a reform here avoiding an ambiguity (or at least a formal defect) in the utterance classes which we will describe below.

The Loglan logically connected sentence is a unit sentence optionally followed by an ICA connective followed by a logically connected sentence: this is a chain of unit sentences linked with ICA connectives. These are understood to group to the left.

The Loglan sentence is then either a logically connected sentence or a sentence with head terms (class **uttAx**): this is a term list, followed by **gi** or **goi**, followed by a logically connected sentence, optionally closed with the right closer GIUO (either **giuo** or **gu**). A head term list marked with **goi** is a prenex quantifier string: the semantics of this needs to be discussed.

Attaching initial modifiers with **gi** is recommended as a matter of style (it is easy to tell where the modifier ends), but with due attention to the fact that such a modifier applies to all unit sentences in a logically connected sentence. This is usually preferred to attaching a modifier directly to the front of a statement or imperative.

A term list attached with **gi** supplies the same final arguments to each unit sentence in the logically connected sentence to which it is attached. The original proposal in 1989 Loglan is that these will be the *last* arguments of the predicate in each sentence. We regard this as a seriously bad idea and have a different proposal. We would suggest that by default the fronted arguments are the *next* available arguments of the predicate (when untagged), and that when an untagged argument follows a tagged argument in the fronted term list it is read as the next argument after the position of the previous tagged argument, if possible (if the tagged argument is not itself last). Note that this allows implementation of the sentence component orders OSV and OVS in which the object is first, with the caution that one again has to attend to the fact that any fronted term list applies to all sentences in the logically connected sentence to which it is fronted.

Provision of a special right closer **giuo** for sentences with head terms is a new proposal: we believe that these may actually see use.

Quantifier Scope

This seems to be the correct point in the grammar to give the precise definition of Loglan quantifier scope. The scope of an indefinite whose scope is not explicitly given by a prefix with **goi** will be the smallest component sentence (unit sentence, logically connected sentence or sentence with head terms) of the utterance which can be its scope. All occurrences of a given indefinite or pronouns referring to it in the same sentence have the same reference: one then chooses from this sentence as the scope the smallest subsentence containing all occurrences of the given indefinite, and all occurrences of indefinites which follow it anywhere within its scope (this last condition may force the scope to be larger, and indeed may force it to become yet larger due to the second identifier being followed by a third one in the absence of the first, and so forth).

Where two indefinites have the same scope, the outer quantifier is determined by which one appears first in the utterance. **Raba cluva be** means that for every x there is a y such that x loves y . **Be nu cluva raba** means that there is a y such that for every x , x loves y (notice how use of the converse enables us to draw this distinction, much as it can in English). There is a qualification: where the apparent first occurrence of an indefinite has an appended subordinate clause (with **JI** or **JIO**) which contains an occurrence of a pronoun referring back to the indefinite, the first such pronoun counts as

the first occurrence of the indefinite (this effect is recursive in nested subordinate clauses). **To mrenu jio te fumna ga cluva mei** refers to three men who are loved by *the same* three women, while **To mrenu jio mei nu cluva te fumna** are two men each of whom are loved by three women (possibly different women for each man).

A prenex quantifier string ending in **goi** has as its scope exactly the sentence with head terms for which it is the head term list. Modifiers in the prenex list are just modifiers. The items in the quantifier list must be indefinites. They may be qualified with subordinate clauses with **ji** or **jio** to indicate restricted scope (use of **jio** allows more complex restrictions than the usual logical notation). Components of the form **ra bua** (where **bua** is a predicate bound variable) are read as quantifying over predicates rather than being bounded to an indefinite predicate. **Ra mrenu goi mei cluva mei** quantifies over men (all men love themselves). **Ra bua ra ba ra be goi ba bua be, ico be bua ba** makes the unlikely assertion that all relations are symmetric: it quantifies over all predicates rather than all things which are **bua**. This is a dodge to accommodate second order logic.

4.2.4 Note on right closers

There are various right closer classes, which we will identify as they come up in the grammar. All the right closer classes can take the shape **gu**, and indeed this was originally the only right closer. Use of the other shape of a right closer class (each can be **gu** or a different closer word) will often avert the need for more than one occurrence of **gu**. A right closer may always be optionally preceded by a pause, and may always optionally be followed by a free modifier.

There are a number of new right closers in this proposal, all rarer than the **guo**, **gui**, **gue**, **guu** already familiar from 1989 Loglan, subdividing the 1989 Loglan class gap. Further, the 1989 Loglan right closers **guu**, **guo**, **gui** have had their grammar significantly changed, though their practical uses are essentially the same, in general with the aim of minimizing or eliminatiing any need for multiple closers at the same point.

4.2.5 Free Modifiers

Free modifiers (colloquially, freemods) are grammar elements which can appear in a very wide range of positions in Loglan utterances. Almost every

medial position between items in a Loglan grammatical rule permits insertion of a free modifier. Free modifiers do not appear in initial positions with a single exception for complete Loglan utterances and appear in final positions only after right closers and in rules defining classes which are in some grammatical sense “atomic”. The usual intention is that a free modifier “modifies” what precedes it immediately, if it does not vaguely modify the entire utterance. These are also constructions which are semantically vague and mostly will play no role in analysis of the logical force of a Loglan utterance.

Pauses which are not phonetically mandatory (i.e., those are preceded by a vowel, followed by a consonant, and not followed by a logical connective or a name word) are free modifiers, though they do not have any content.

The other flavors of free modifier are:

1. Words of class UI or NOUI (attitudinals and negative attitudinals).
2. Spoken smilies: **soi** followed by a descriptive predicate (class **descpred**) optionally closed with the right closer **guea** or **gu**.
3. One of the register words of class DIE, optionally negated by being prefixed with **no**.
4. A parenthetical utterance **kie** + utterance + **kiu**. The utterance must be a well formed Loglan utterance. It may be set off (after **kie**, before **kiu**) either with commas or with a pair of parentheses.
5. inverse vocatives, to be discussed just below.
6. vocatives, to be discussed just below.
7. items of class JO, the word **jo** optionally preceded by a digit, which has the effect of putting a number of words preceding the JO item in “scare quotes”. The number of words “quoted” is indicated by the digit if it is present, and otherwise is one.
8. ellipses . . . and double hyphens --.

A freemod can always optionally include another appended freemod.

Vocatives

A vocative is a free modifier indicating who is addressed by the speaker of the utterance. It will begin with a vocative marker, either **hoi** (which we should recall is a name marker) or one of the words of social lubrication, **loi, loa, sia, sie, siu**. The words of social lubrication are not name markers, so pauses before names are required: **Hoi Djan**, but **Loa, Djan**. A free modifier may not occur between a vocative marker and the following utterance: this prevents subjects being grabbed by the social lubrication words when they are used as vocative markers: it enables the old form **Loi hoi Djan** to work. What follows the vocative marker is either

1. a name (possibly set off from the vocative marker by a comma-marked pause),
2. or a descriptive predicate (**descpred**) which may optionally be closed with the right closer GUEA (**guea** or **gu**) optionally followed by a name, which may be marked with **ci** and must be so marked if it contains a false name marker.
3. or an argument without a case tag (**argument1**), possibly set off from the vocative marker by a comma-marked pause and possibly closed with the right closer GUUA (**guua** or **gu**),
4. or a foreign name (alien text): in this case the vocative marker must be **hoi** and the alien text must be enclosed in double quotes. The double quote requirement is to prevent accidental acceptance of buggy Loglan text as a foreign name.

The special closers here are products of a proposed subdivision of the old class gap.

Inverse Vocatives

Inverse vocatives are free modifiers indicating who is uttering the text. They always begin with the inverse vocative marker **hue**, which is a name marker. There is not a restriction on free modifiers at joints as in the vocative (of course there cannot be a free modifier between **hue** and a following name).

1. A name.

2. A descriptive predicate, possibly closed with the right closer GUEA, possibly followed by a name which may be marked with **ci**. This is just as in the vocative construction, except that a free modifier is allowed after **hue**.
3. A statement possibly closed with the right closer GIUO (**giuo** or **gu**). This allows free modifiers of the form “said John”, as it were.
4. An argument without a case tag (**argument1**), possibly closed with the right closer **guu** or **gu**.
5. A foreign name (alien text), which must be double quoted.

4.2.6 Utterances

We will refer to elements of the class of utterances **uttA1** as “general answers”. A general answer is one of the following:

1. A unit sentence.
2. A sentence with head terms (notice that a mere logically connected sentence is not of this class).
3. tightly linked term lists with **je** and **jue** (classes **links** and **linkargs**, described below).
4. subordinate clauses (class **argmod**).
5. a term list. NOTE: why are termsets not in this class?
6. a brief answer of class **uttA** (either a logical connective or a quantifier/number).
7. An occurrence of **no** as a word.

A general answer may optionally be suffixed with terminal punctuation.

An utterance of class **uttC** is either a general answer or a sequence of negheads followed by a general answer.

An utterance of class **uttD** is one of two things:

1. A logically connected sentence, optionally followed by terminal punctuation, not followed by an ICI or ICA connective

2. A sequence of **uttC** utterances linked with ICI connectives.

An utterance of class **uttE** is an utterance of class **uttD** or a chain of such utterances linked with ICA connectives.

Notice that we have arranged for a logically connected sentence such as **da redro, ice da blanu**

to be parsed as a single **uttD** unit (since it is a logically connected sentence), rather than as two **uttD** units linked by an ICA connective, which is how previous grammars would have parsed it if it appeared by itself as an utterance (which is downright weird, as it also admits a parse as a logically connected sentence, which would occur in other contexts). I do not know if this was technically an ambiguity in previous grammars, but it was certainly a formal defect.

An utterance of class **uttF** is a single **uttE** unit or a chain of **uttE** units linked with class I utterance connectives (which include IPA or IKOU connectives such as **irau**).

A Loglan utterance is one of the following (with the side condition that it cannot begin with the **cmapua ge**):

1. an I connective, optionally followed by terminal punctuation and optionally further followed by an I-connective initial utterance.
2. a free modifier other than a pause (optionally prefixed with an I connective and optionally suffixed with terminal punctuation, and further optionally followed by an utterance of any form). This is essentially the only case where a free modifier might appear initially.
3. an **uttF** followed by an IGE connective followed by an utterance.
4. an **uttF** optionally prefixed with an I- or ICA connective, followed optionally by an I connective-initial utterance.

There are two ways in which an utterance can occur. It is either top level, and so followed by end of text (which does include weird options with **#**), or occurs as part of a **li-lu** quotation or a **kie-kiu** parenthetical remark.

We have climbed as far up the parse tree as we can and now must climb downward.

4.3 Predicates

Our treatment of predicates may conveniently be divided into treatment of verb phrases (which do not have termsets attached, though component predunits may have linked term sets built with **je** and **jue** attached) and predicates per se, which may have termsets attached. We note that the word predicate has two uses: here it is being used to describe the part of a sentence consisting of a verb phrase and objects, whereas in earlier contexts it is used to describe predicate words. We believe that the reader should be able to distinguish these usages.

4.3.1 Verb Phrases

We begin with simple verb phrases (which are not all that simple) then proceed to complex verb phrases constructed by processes of adverbial modification (“metaphor”) and logical connection with CA family connectives.

Simple Verb Phrases

The ultimate building block of verb phrases is the class **predunit1**. A **predunit1** is one of the following:

1. A predicate word, possibly preceded by a conversion or reflexive operator of class NU.
2. A foreign predicate or onomatopoeic predicate (**sue** or **sao** followed by alien text).
3. A possibly complex descriptive predicate turned into a **predunit** using the initial marker **ge** and (optionally) the final marker **geu** or **cue** (**cue** is an older form of class GEU: this is not a right closer class and does not have **gu** as a possible shape). This takes two forms: the basic form is **ge** + **descpred** + (optionally) GEU and the converse or reflexive form NU + **ge** + **despredE** + (optionally) GEU. I need to look into the reasons why slightly different classes of descriptive predicates are allowed here. Grouping with **ge...**(GEU) is useful in expressing complex metaphors precisely.
4. A predicate built from an argument without a case tag: **me** + **argument1** possibly closed with the right closer MEU (**meu** or **gu**).

5. An abstraction predicate: this is one of the abstractors **po**, **pu**, **zo** followed by a sentence (unit sentence, logically connected sentence or sentence with head terms) possibly closed with the right closer **GUO**. Alternatively, forms of the abstractors suffixed with **(z)a**, **e**, **(z)i**, **o**, **(z)u** may be used, in which case forms of the right closer with the same suffix may be used: this allows closure of several nested abstraction predicates (or abstract descriptions) with a single right closer.

In the trial 85 grammar these predicates could only occur at the very top level of the predicate parse tree; they could not, for example, participate in metaphors. Their present position in the grammar makes much more sense.

Free modifiers are allowed in medial positions in these constructions (except between **sao/sue** and alien text). I specifically allow pauses before **ge** and after **geu**. Moreover, this is one of the few rules which allows an optional free modifier in final position (such classes are “atomic” in some sense; the right closer classes also have this characteristic).

A **predunit2** is a **predunit1** possibly preceded by one or more occurrences of **no**. **No** binds very tightly to predunits initial in metaphors; to negate a verb phrase or predicate may require some initial marking to avoid the **no** being absorbed into a **predunit2** instead. **No kukra prano** means “to run slowly” (**prano** modified by **no kukra**): note that this asserts that you run, though not quickly. **no ga kukra prano** means “not to run fast”; this does not say you run at all.

A **predunit3** is a **predunit2** possibly followed by a linked term set built with **je/jue**.

Finally, a predunit is either a **predunit3** or a **predunit3** preceded by a short-scope event abstractor, one of **poi**, **pui**, **zoi**. These replace the short scope uses of the original abstractors in 1989 Loglan: all occurrences of **po**, **pu**, **zo** are long scope. The predunit class is of particular note because it is the sort of predicate which can occur as a component in a serial name.

A further “unit” predicate is the forethought connected predicate, which consists of an optional prefix of one or more **no**’s, followed by a forethought connective (**KA**) followed by a predicate (of the most general form), followed by a class **KI** word completing the forethought connection, followed by another predicate, optionally closed with the right closer **guu** or **gu**. Note that the **guu** closure, introduced late, removed the reasons for the rule forbidding forethought connected predicates as heads of adverbial modifications

(“metaphors”). This is an arbitrarily complex predicate construction which is treated as a simple verb phrase because it is suitably packaged.

Again, free modifiers are allowed in medial position in all grammar constructions described here (except after **sao/sue**).

Complex verb phrases formed with adverbial modification (“metaphor”) and tight logical connection

The construction of complex verb phrases by a combination of adverbial modification and logical connection is the subject of this little section.

I am referring to the process which JCB calls “metaphor” as “adverbial modification”.

A **despredA** is a series of one or more predunits or forethought connected verb phrases separated by the little word **ci**. There are occasional phonetic issues caused by the fact that **ci** is a name marker, though that is quite irrelevant to this particular use of the word. This is a adverbial modification construction, and it groups to the left.

A **despredC** is a series of **despredB**’s with no intervening structure word: this is a adverbial modification construction, binding more loosely than the construction with **ci**. This particular construction is actually used only internally to **despredB**.

A **despredB** is either a **despredA** (the usual situation) or a construction of the form (**cui** + **despredC** + **CA** + **despredB**). The idea is that arbitrarily many **despredC**’s can be logically connected on the left to a **despredA**, the left boundaries of the **despredC**’s being guarded with **cui**.

A **despredD** is a series of **despredB**’s separated by **CA** series logical connectives: these are logically connected verb phrases, grouping to the left.

A **despredE** is a series of **despredD**’s without intervening operators: this is the general purpose adverbial modification construction, grouping to the left. Right grouping can be forced using **ge...**(**GEU**). (NOTE: this is the class which can be packaged with **ge** and converted in **predunit1**: I do not know why the restriction to **despredE** is imposed and may allow **despred** there as well.)

A descriptive predicate (**descpred**) is either a **despredE** or a **despredE** followed by **go** followed by a descriptive predicate. The latter case is inverse adverbial modification: the descriptive predicate after the **go** modifies the initial **despredE**. This class has many uses (and thus an English name in our grammar). Note that in our terminology a descriptive predicate might better

be called a descriptive verb phrase.

A sentence predicate (or sentence verb phrase) is either a **despredE** or a **despredE** followed by **go** followed by a **barepred** (a class of sentence predicates in the proper sense described below). The difference is that this construction may end with a termset, but it is still a verb phrase, as the termset is attached to a subordinate part of the phrase. Our grammar differs here from the trial 85 grammar in not drawing a systematic difference at all levels between sentence and description verb phrases, which simply turned out not to be necessary, once the restriction on forethought connected predicates in head position in predicate modifications was removed.

Free modifiers are allowed in medial positions in all constructions described here.

4.3.2 Predicates

We now discuss predicate constructions involving termsets.

The most basic of these is the bare predicate (class **barepred**) which consists of a verb phrase, optionally followed either by a termset (a black box concept for now: it will be explained below) or by the right closer **GUU** (**guu** or **gu**) if this is in turn followed by another termset. The **GUU** class was originally conceived as a closer for termsets (and **GUU** by itself was an empty term set). Defining **GUU** as a closer of bare predicates, used only to separate a bare predicate from a following termset not attached to it, has very similar effects in practice and very often allows fewer closers to be used.

A marked predicate (class **markpred**) is a bare predicate prefixed with a **PA** class word (a tense, in the most general sense) or the null tense **ga**. This is as good a place as any to note that we do not endorse JCB's definition of the semantics of **ga** as explicitly asserting potentiality. We view it as simply noncommittal (with the reading of the sentence as expressing potentiality often a reasonable reading).

A **backpred1** is a bare predicate or marked predicate optionally prefixed with one or more occurrences of **no** (where these occurrences of **no** are not absorbed into words or initial predicates in adverbial modification constructions; inserting **ga** can avert such problems in negating a bare predicate).

A **backpred** is a **backpred1** by itself, or a sequence of **backpred1**'s separated by **ACI** logical connectives, followed optionally either by a termset (optionally closed with **GUU**) or **GUU** by itself (only if followed by a termset not in the **backpred**), further followed optionally by one or more units con-

sisting of an ACI connective followed by a **backpred**, followed optionally (again) by a termset (optionally closed with GUU) or GUU by itself (only if followed by a termset not in the **backpred**). The termsets following logically linked **backpred1s** or **backpreds** are shared by all the logically linked **backpreds**, in both cases. (arguments in the second appended termset are supplied to the verbs which received the first appended termset, as well).

A **predicate2** is a **backpred** by itself (not beginning with **ge**), or a sequence of **backpred**'s (not beginning with **ge**) separated by A logical connectives, followed optionally either by a termset (optionally closed with GUU) or GUU by itself (only if followed by a termset not in the **backpred**), further followed optionally by one or more units consisting of an A connective followed by a **predicate2**, followed optionally (again) by a termset (optionally closed with GUU) or GUU by itself (only if followed by a termset not in the **predicate2**). The termsets following logically linked **backpreds** or **predicate2s** are shared by all the logically linked **backpreds**, in both cases (arguments in the second appended termset are supplied to the verbs which received the first appended termset, as well).

This description is rather baroque: it is actually driven in its structure by limitations of PEG grammars. The solution to appending shared final termsets to logically connected predicates in the *trial.85* grammar is very elegant but hopelessly left recursive in a way PEGs cannot manage. It is also to be noted that in *trial.85* the ACI connectives are not really fully usable logical connectives: here they are a fully privileged sequence of logical connectives binding more tightly than the A connectives. It should be noted that the A and ACI connectives group to the left. We believe that all patterns of logical connection and sharing of final termsets which could actually be spoken in practice are supported. Note again that GUU does not act as a right closer for termsets (it is not included in the termset as a final component) as it was in *trial.85*, but serves to terminate a termset in the context of a larger class, or protect an instance of the larger class from absorbing a following termset which it should not include.³

A **predicate1** is either a **predicate2** or a **predicate2** followed by an AGE logical connective, followed by a **predicate1**. These logical connectives group to the right and bind most loosely of all. The restriction on **predicate2s** starting with **ge** is both harmless (this is not normally something one needs in initial position) and required to avoid a phonetic ambiguity

³This is an obvious place where support by examples is required.

if an A connective is followed by a **ge** initial predicate, which was not noted by our Founders.

An **identpred** is an identity predicate (class BI, see the lexicography section) optionally preceded by a finite number of occurrences of **no**.

A predicate is either a **predicate1** or an **identpred**.

Notice that the grammatical privileges of identity predicates are quite limited.

Free modifiers can appear in all medial locations in the grammar rules given here.

This completes the grammatical account of predicates!

4.4 Arguments, Modifiers, and Term Lists

In this section we handle the noun phrase (and relative clause) side of things.

4.4.1 Predicate Modifiers

We begin with the relatively easy description of *predicate modifiers* (relative clauses). These function to some extent like additional arguments in a sentence, and they are grouped with the arguments in a general class of “terms”.

A *mod1* is either (1) a PA word followed by an argument which is not case tagged, optionally followed by the right closer GUUA (**guua** or **gu**) or (2) a PA phrase not followed by a bare predicate (so it is not a tense) optionally closed with **gu**. The more specific closer GUUA for relative clauses is a new proposal.

A *kekmod* is a expression KA + modifier + KI + mod, optionally prefixed with one or more negations. This is forethought logical connection of modifiers.

A *mod* is a *mod1*, possibly multiply negated, or a *kekmod*.

A modifier is a *mod* or a sequence of mods separated by A connectives.

4.4.2 Serial names

We describe the full construction of complex proper names.

A complex proper name always begins with a name word (with or without false name markers). Subsequent units are of the following shapes:

1. A name word not containing a false name marker.
2. A name (including acronymic names) marked initially with **ci**.
3. A predunit, marked initially with **ci**, which cannot be followed by a name word without name marker. This allows forms like **la Djan ci Blanu**, “John the Blue”. Note that forms like **la Bilti, Djin** are not serial names (they are a kind of description).

Pauses are required between the units in a serial name, but they do not have to be written explicitly with commas. The proposal in Appendix H of two grades of pause (one occurring only in serial names) has been abandoned:

it became unnecessary when **la Djan Blanu** ceased to be a name and could no longer be confused with a simple sentence.

A pause at the end of a serial name is not actually required if the last unit in it is a predunit, but it is strongly advisable to pause at the end anyway, as **ci** is a name marker and untoward effects might otherwise occur.

4.4.3 Arguments (and subordinate clauses)

A first variety of argument to mention is the class LANAME, which is simply **la** followed by a serial name, with an optional intervening pause.

A description (class **descriptn**) is one of a number of forms, the most common of which begin with articles of class LE.

1. unordered lists with **lau** or ordered lists with **lou**: these arguments were apparently in an early stage of development in trial.85 and had very limited grammatical privileges. We have restricted and modified the form of these terms. They also had a quite alarming freedom of form. Unordered lists now start with **lau**, are followed by one or more arguments of the forms **arg1a** or **indef2**, separated by commas **zeia** if there are more than one, and closed with **lua**. The ordered lists are the same, with opener **lou**, comma **zeio**, and closer **luo**. These are spoken analogues of the usual finite list notation for lists or sets. NOTE: I should move these to be a case in **arg1a**.
2. **ge** followed by a quantifier followed by a descriptive predicate. This is semantically an indefinite: we want to investigate contexts in which this form needs to be used. NOTE: I should move these to be a case in **arg1a**, unless I really want the option of affixing a name to them. This might actually be nice: **ge to Nirli, Smit**, “two Misses Smith”.
3. Forms beginning with class **LE** articles: if these begin with **la**, they must not be of class LANAME. The basic form is LE + (optional argument of class **arg1a**, not beginning with a quantifier) + (optional tense (PA phrase)) + (optional quantifier)+ descriptive predicate. A classic example is **le mrenu**. The optional components may be seen in **lemi hasfa** or **le la Djan, hasfa**, and in **levi hasfa, lemina hasfa** and even **le la Djan, na hasfa**. The optional argument component is a “possessive”; the tense or location component has obvious effects, and the quantifier has obvious effects, as in **le to mrenu**. An example

with all three components is **le la Djan, na to hasfa**, “John’s two present houses”. A variant form is LE + (optional argument of class **arg1a**, not beginning with a quantifier) + (optional tense (PA phrase)) + quantifier + argument of class **arg1a**, as in **Le to le mrenu** “The two of the men”.

There is a change here from 1989 Loglan which is perhaps worth noting. In 1989 Loglan (and in early versions of my parser) **lemi hasfa** was not of the same grammatical form as **le la Djan, hasfa**, though surely they felt parallel. **lemi** was originally a word, and the grammar rule enabling **le la Djan, hasfa** did not accept a pronoun in the possessive position. We made these forms grammatically parallel as part of a process of eliminating multisyllable *cmapua* where this could be done in a conservative way. The form **le, la Djan, na hasfa** then became legal because of the parallelism with **lemina hasfa**.

The next special form considered is the abstract description. This consists of a word of class LEFORPO (either **le**, **lo**, or a quantifier core (class NI2)) followed by a word of class PO (**po**, **pu**, **zo**, depending on whether an event, property, or quantity abstraction is being formed), followed by a unit sentence, logically connected sentence, or sentence with head terms, followed optionally by a right closer of class GUO (**guo** or **gu**). To facilitate closing multiple abstractions at once, one may suffix the PO word with one of **(z)a**, **e**, **(z)i**, **o**, **(z)u**, in which case the closer (if present) must be **guo** with the same suffix. The suffixed closers are a NEW proposal.

It is important to note that **le po mrenu** does not have an abstraction predicate **po mrenu** as a grammatical component, which averts the need for a double closure of such arguments apparently found in Lojban. There is an occasional need for a **ge** when an abstraction predicate really is intended after a LEFORPO article, as in **Le ge po sucmi guo ditca**, a teacher of events of swimming, as opposed to **Le po sucmi guo ditca**, a sentence asserting that the event of swimming teaches!

The classic **le, po** issue found in 1989 Loglan no longer exists. **lepo sucmi ditca** and **le, po sucmi ditca** both mean “the swimming lesson”. **le poi sucmi ditca**, using the new short scope form, is “the swimming teacher”. One could also say **le po sucmi guo ditca** for “the swimming teacher”, but why? **lepo**, by the way, is not a word.

We now describe a general class of “atomic” arguments **arg1a**. This is a class which can be followed directly by an optional freemod. An argument

of this class is one of the following:

1. A pronoun (class DA or TAI).
2. An abstract description.
3. A numerical description: **lio** followed by a descriptive predicate and optionally a right closer (**guea** or **gu**), or an argument without case tag and optionally a right closer (**guua** or **gu**), or a quantifier (**mex**) followed optionally by the right closer **gu**, or alien text (NOTE: have I listed **lio** as a possible alien text marker? Double quotes might reasonably be required here: actually, this case is inoperative, because it will fail phonetics checks until I make **lio** an alien text marker; I need to fix this, because it is wanted for numerals)
4. a foreign name with **lao**, already described in the phonetics section.
5. a **LANAME**, always read in preference to a description.
6. a description, optionally closed with the right closer **guua** or **gu** and further possibly followed by a name, which can be marked with **ci** and must be so marked if it contains a false name marker.
7. a strong quotation with **lie**, quoted word with **liu**, or quoted Loglan text with **li...lu**, all described above.
8. **ge** followed by an **arg1a**.

Freemods can appear in all medial positions in these constructions, except for known restrictions on names, alien text, and quotation forms, and can also appear in final position because this is an “atomic” argument.

We now describe the formation of subordinate clauses with **ji**, **ja**, **jio** or **jao**. The intention of **ji** and **jio** is restrictive, supplying additional information about the reference of the identifier to which the subordinate clause is attached: **le mrenu ji vi skitu**, “the man sitting here”, versus **le mrenu ja vi skitu**, “the man whose identity is already understood, who happens to be sitting here”. The form with **ji** tells us who we are talking about: the form with **ja** supplies additional information. The distinction between **jio** and **jao** is similar.

A basic subordinate clause with **ji** or **ja** consists of the word followed by a predicate or a term (modifier or argument without case tag). A basic

subordinate clause with **jio** or **jao** consists of the word followed by a unit sentence, logically connected sentence, or sentence with head terms. Any of these forms may optionally be prefixed with **no**. If the word of class JI or JIO is suffixed with one of **-za**, **-zi**, **-zu** then the basic subordinate clause may optionally be closed with **gui** followed by the same suffix.

A subordinate clause is either a basic subordinate clause or a series of basic subordinate clauses inked by A series logical connectives, and may optionally be closed with the right closer **gui** or **gu**.

The use of **gui** to close logically linked sets of subordinate clauses and the use of special forms to close basic subordinate clauses are improvements to the original device of closing just basic subordinate clauses with **gui**. As is generally the case with our adjustments of right closers, the intention is to avoid as far as possible ever having to utter two right closers in practice.

An argument of class **arg2** is an **arg2a** class argument optionally followed by one or more subordinate clauses.

An argument of class **arg3** is either an argument of class **arg2** or a quantifier followed by an argument of class **arg2**: **to le mrenu**, “two of the men”.

An argument of class **indef1** is a quantifier followed by a descriptive predicate: **to mrenu**, “two men”.

An indefinite argument is an argument of class **indef1**, possibly closed with the right closer **guua** or **gu**, followed optionally by one or more subordinate clauses.

An argument of class **arg4** is an **arg3** or indefinite argument or a sequence of such arguments linked by the fusion connective **ze**.

An argument of class **arg5** is either of class **arg4** or of the form KA + argument without case tag + KI + **argx** (which is explained next). In other words, these are forethought connected arguments, the last being **argx**, which is basically an **arg5** with additional detail allowed which is detailed below.

An argument of class **argx** is (optional **no**) + (optional LAE argument of indirect reference, or more than one) + an **arg5** argument.

An argument of class **arg7** is an **argx** or a sequence of **argx**'s linked by ACI series logical connectives. These group to the left.

An argument of class **arg8** may not begin with the cmapua **ge** for technical reasons, and is otherwise an **arg7** or a sequence of **arg7**'s linked by A series logical connectives. These group to the left.

An argument of class **argument1** (an argument without a case tag) is an **arg8** argument, followed optionally by an AGE connective followed by an

argument without a case tag. The AGE connectives group to the right. This can further optionally be followed by **guu** (not **gu**) followed by a subordinate clause (allowing attachment of a subordinate clause to complex arguments).

An argument is of the form (optional one or more **no**'s) + (optional one or more case tags (DIO class words))+ an argument of class **argument1**.

Free modifiers are allowed in all medial positions in these constructions.

4.4.4 Term lists and termsets

A term is an argument or predicate modifier (relative clause).

Class **terms** (term lists) is inhabited by concatenations of one or more terms, containing no more than four un-case-tagged arguments. The four un-tagged arguments are parsed with classes **argumentA**, **argumentB**, **argumentC**, **argumentD** in that order. In the usual parser, the lettered classes are actually just the same as **argument** and the class names are just a hint about argument places.

In the alternative parser, the further restriction is placed on the classes **argumentA**, **argumentB**, **argumentC**, **argumentD** that an argument of one of these classes cannot stand at the beginning of a statement which is not marked with the large subject marker **gaa**. The idea is that this guards term lists from picking up the subject of a following sentence. A sentence like **Na lepo la Djan, bleka mi, mi bleka la Djan** (While John looks at me, I look at John) will be read correctly without the need to close the **lepo** clause with **guo**, and without ascribing any significance to the pause after the initial clause. Under the alternative parser, it will sometimes be necessary to follow a large subject with trailing arguments (which might come from subordinate clauses with **JI/JIO** as well as from **lepo**-style abstract descriptions) with the large subject marker, so that it will not read a trailing argument as the subject of a following sentence.

Under the alternative parser, there is another class of term lists using alternative classes **argumentA1**, **argumentB1**, **argumentC1**, **argumentD1** which do not have the restriction that they cannot stand at the beginning of a statement. This class of term lists is used after **gio**, where this restriction would be inappropriate.

Class **modifiers** is inhabited by term lists containing only modifiers.

Class **modifiersx** is inhabited by term lists in which any modifiers are case tagged.

Class **subject** is inhabited by concatenations of terms which contain at least one argument and at most one argument without a case tag: the untagged argument is not assigned class **argumentA**: a subject is not a species of the class **term**.

We now discuss termsets.

Class **termset1** is inhabited by term lists of class **terms**, and by structures of the form **KA** + **termset2** + (optional **guu** or **gu**) + **KI** + **termset1**, forethought connected termsets.

Class **termset2** consists of one or more **termset1**'s connected by **A** connectives, each **termset1** being optionally followed by the right closer **guu** or **gu** if this is itself followed by an **A** connective (not necessarily in the **termset2**). The idea of the closure is to insulate logical connectives connecting termsets from being interpreted as logical connectives connecting arguments. Notice again that the **GUU** right closer is not part of the **termset1** being closed, but part of the ambient **termset2**. I think that in this case the **GUU** will always express itself before **A** connectives linking component **termset1**'s, and there may or may not be an additional one after the last component **termset1**, separating the **termset2** from a following **A** connective not included in it.

A termset is either a **termset2** or a term list (class **terms**) followed by **go** followed by a bare predicate. The last construction is entertaining: the final bare predicate will modify (as a adverbial modifier) the verb phrase to which the termset is attached. NOTE: why does this last construction only allow a term list and not more complicated termsets?

We have seen the roles of the various flavors of term lists and termsets in other grammatical constructions above.

4.4.5 Linked term sets with **je/jue**

We close by discussing linked term sets. We have seen above that these attach very tightly to predunits.

In general terms, these are term lists in which the first term is attached to the predunit by **je** and subsequent terms are attached with **jue**. There are technical details, laid out in what follows.

A **jelink** is **je** followed by a term or a **PA** phrase (class **PA2**). If what follows is a **PA2**, it may optionally be followed with **gu**.

A **juelink** is as a **jelink** but with **jue** instead of **je**.

Allowing modifiers as well as arguments is new in TLI Loglan: my understanding is that it was already allowed in Lojban.

A **links1** is a sequence of one or more **juelinks** optionally followed by the right closer **gue** or **gu**.

A **links** is a **links1** or a structure KA + links + KI + links1, either of these alternatives further being optionally followed by additional **links1**'s connected by A connectives. NOTE: there is something odd here about the restrictions on use of forethought connectives, but how often will this be done in any case? Testing is a good idea.

A **linkargs1** is a **jelink** followed optionally either by a **links** or by **gue** or **gu**. Note that any need for a double closure is averted by requiring that it be one or the other.

A linked term set (class **linkargs**) is defined exactly as a **links** is, using **linkargs1/linkargs** in place of **links1/links**.

4.4.6 The Alternative Parser, summarized

I do propose adoption of the “alternative parser” approach to certain closures. In this section, we describe the grammar changes involved in adoption of this further proposal in one place.

The basic idea is that a term list **terms** will not pick up a term which it recognizes as the subject of a following sentence (term lists following **gio** do not have this restriction). This is achieved by restricting the classes **argumentA**, **argumentB**, **argumentC**, **argumentD** to not stand at the head of a statement unless the statement is marked with the “large subject marker” **gaa**. The fact that the grammar change is executed by a change in the definition of term lists is the reason why this section appears where it does.

Notice that the large subject marker must be stressed, and if it is finally stressed, one must pause after it if it is followed by a predicate word. These are virtues. It may resemble **ga** phonetically and often replace it, but it is not a tense and can be followed by a tense (or even by **ga**). The large subject marker may appear after the subject in a statement or initially in a gasent or imperative, as described above.

In **Na lepo la Djan, bleka mi, mi bleka la Djan**, the alternative parser recognizes that the second **mi** should not be a third argument of **bleka**, so this is successfully read as “While John watches me, I watch John”, and not by use of pause/**gu** equivalence. The pause is stylistically a good idea, but it is not required for the sentence to parse as desired.

Na lepo la Djan bleka, tu prano means, “While John looks, you run”. **Na lepo la Djan bleka tu gaa, prano!** means “When John looks at you, run!”: it is an imperative, and the **tu** is guarded from being taken as the subject of a statement by the occurrence of **gaa**. The pauses are purely decorative and have no effect on the grammar.

In **Lepo tu cluva mi gaa gudbi**, the use of **gaa** prevents **mi** from being rejected as an argument of **cluva**.

In **Le mrenu ji cluva mi gaa fia sadji**, the use of **gaa** is also required: trailing arguments may occur in subordinate clauses as well as in abstract descriptions or predicates. We also illustrate use of a tense after **gaa**.

This device allows many complex constructions which occur initially in a statement or are followed by a statement, gasent, or imperative to be closed efficiently without the use of one or more right closers which would otherwise be needed. We think the situations in which **gaa** is needed are recognizable (it marks large subjects with trailing arguments which might be misinterpreted as subjects to the following predicate, and it is uttered after they are uttered and so can be added as an afterthought). We do not simply adopt this as part of our grand proposal, because we regard it as more of a departure than our other changes, though in a certain sense it is conservative in allowing recovery of many constructions justified by pause/**gu** equivalence in earlier versions of the language.