Draft Loglan Reference Grammar

Randall Holmes

September 12, 2015

Contents

1 Introduction

The purpose of this document is to give an independent description of the TLI Loglan language, in the provisional version embodied in my PEG parser, without PEG notations which are difficult for a nontechnical reader to follow. This does not mean that technicalities do not arise. One purpose of this is to give a clearly independent description of my intentions which can be used to double check the PEG parser. Another purpose is to give a venue for presenting material which is not in the purview of the parser, such as stating the semantics of grammatical words which are just items in lists for the parser.

An important point is that a lot of jargon (names of grammatical classes) is needed, paralleling structures in the PEG grammar and indeed in trial.85. I would like to create Loglan predicates for many or even all of these terms, first

reducing their use as much as possible. There is a side project of creating an adequate native grammatical vocabulary.

In intention, this document is to be a complete description of the language. It does not represent a power grab on my part: this is to be modified in accordance with any decision of *la Keugru*. As always, I am well aware that my parser is not yet official. But I am not modest about the fact that I have definite ideas about how things will go, and I feel free to put them in here – but also obligated to point out proposals implicit in this text. NOTE: 7/16/2015 I need to be sure to point out all places where 1989 Loglan has been modified, with or without an official academy decision supporting it.

My intentions are conservative. I do not feel committed to not making changes as I work on this but I am committed to the language described being intelligible to a speaker of 1989 Loglan (if such a being existed).

The descriptions given in the phonology section are often far simpler than the PEG code makes them look. Some of the specifications are quite awkward to achieve with a PEG (and would be even more awkward with a BNF grammar).

As of 7/16/2015 this document is finished at a zero draft level, except that I have quite deliberately omitted explicit set and ordered list forms, as I think their grammar needs to be rethought. The way they were slotted into the trial.85 grammar is exceedingly awkward. The 7/16/2015 text includes notes to myself about places where I think the grammar may need minor fixes.

I note that a section on word building from djifoa is needed. The well-formedness of complex predicates is handled in the phonetics section, but the semantics and the pragmatics of building complexes need a venue.

7/16/2015 I have added the full grammar section, at the level of a zero draft. This text is fairly complete on the level of grammar, but needs lots more discussion of semantics, as well as examples and motivation and discussion of the pragmatics of using the grammar to say things we want to say.

3/28/2015 added an important statement about the concept of "word" to the foreword of the lexicography section and corrected some misleading statements in the phonetics section.

5/15/2015 global revisions. Added comments reflecting 5/15 revision of parser which forbids spaces internal to words.

5/16/2015 refinments of the 5/15 revision ruling out spaces inside of words. Removed capitalization restrictions in acronymic names, so `la DaiNaizA` is legal, for example.

5/20/2015 noticed that **ii** or **uu** as an explicit disyllable should be included in RepeatedVowel. This does rule out some forms for predicates with explicit syllable breaks in **ii** or **uu** diphthongs.

5/22/2015 an editing pass

6/3/2015 starting the Grammar section. Wrote a section on sentences. I am going to work more or less backward in the formal grammar and `loglangenda`, as this seems to promote a more natural top down approach to the material.

6/4/2015 added the section on free modifiers and utterances. Do be aware that all of this is very much a draft in terms of style and arrangement of content.

9/5/2015 implemented changes in marked predicate initial sentences.

9/12/2015 further changes to allow sentences with no terms or only modifiers ahead of an unmarked predicate to be recognized as imperatives. It is useful to note that the effect of the 9/5 and 9/12 modifications affects parsability of hardly any sentences. It forbids `terms gasent` sentences with an argument among the initial terms, which JCB said in NB3 make no sense anyway. The main effect is that it reclassifies which sentences are imperative. A sentence with no terms or only modifiers before an unmarked predicate is an imperative; a sentence with no terms or only modifiers before a marked predicate is a declarative sentence with unexpressed subject.

2 Differences from earlier implementations of Loglan

I list some points about the approach I have taken to implementing the phonology/orthography, lexicology and grammar of Loglan. This section is intended only for readers with prior technical knowledge of the language.

general intentions: My intentions have been conservative. My intention is that the language here be intelligible to a speaker of 1989 Loglan if there were such a being. I have made changes and clarifications mostly where I thought some clarification or repair was needed. There is no intention of radical redesign of the language, nor any intention of addition of major improvements at this time.

the perceived difficulty with LIP and trial.85: The official definition of the language was embodied in the trial.85 formal language, the LIP interactive parser. and the document Notebook 3 (NB3) as superseded by 1989 Loglan 1 and further decisions of the academy. There are serious difficulties with this. There is a fairly good description of the phonetics in the sources, but no adequate formal description of the lexicography. The definitions of classes of structure words are implicit in non-human-readable tables in LIP, which evidently have bugs in them. The fact that the lexicography (and the phonetics) are not actually integrated with the formal grammar means that there are limits to any provable unambiguity (and indeed the old definition has demonstrable ambiguities caused by interaction of word classes with the grammar). There are specific bad features of the old definition which I will bring up under other points.

PEG rather than BNF: I chose to write a phrase expression grammar (PEG) rather than a BNF grammar. Issues of ambiguity do not arise with a PEG, as it assigns priorities to alternative parses and picks the first one that works. The hazard analogous to ambiguity in a PEG implementation is "unintended parse", when priorities are set incorrectly. The PEG formalism is very expressive but does act a bit weirdly.

a unified approach to all levels: I wrote a single parser working from the level of letters upward, handling phonology/orthography, lexicography,

and grammar all in the same parser, with no preprocessing. I thought initially that I would eventually write a different parser for Loglan phonetics, but in fact the initial PEG parser developed into a parser capable of handling a phonetic notation as well as traditional Loglan orthography.

phonetics: I believe the implementation of the phonetics is almost exactly as in the original definition. I based my work on borrowed predicates and names on a precise definition of the Loglan syllable. I do require that names resolve into Loglan syllables (which does not affect any names appearing in the corpus). Because of this, I require that names with **la** be written as pronounced, while names with foreign spelling may be written with the **lao** form for foreign names. Early in this work I proposed to the Academy that doubled vowels which force stress not be allowed in borrowings, and this was approved. I require that syllabic consonants be doubled, which does affect spellings of several names in the corpus, but which is also explicitly suggested in Loglan1. I adopted a different rule for grouping long strings of vowels in names or borrowings than is given in the sources. I proposed and the Academy accepted a clarification of the phonetic maneuver that abolished the **slinkui** test, but it appears that the modification may have been a restatement of the original Academy's actual intentions, as the change did not materially affect the dictionary.

phonetic parsing: The development of parsing of predicates required rules that analyzed implicit stress on syllables (which in turn required me to define syllables). It turned out to be fairly easy to adjoin explicit notation for syllable breaks (using the hyphen instead of the close comma) and for stress and emphatic stress on syllables. I then made an effort to ensure that wherever an actual pause is permitted, a comma pause is legal to write (this was not true in LIP). This makes it possible to write in a "phonetic transcript" style in which all stresses of interest are shown (or all stresses and syllable breaks if one wants to be thorough) and all pauses in speech are represented by actual commas, which is handled by the parser as readily as traditional Loglan orthography (and mixed styles are also possible – phonetic notation of stress can be used for emphasis in writing).

A converse feature of the current parser (defending the concept of "word") is that one cannot write a space internal to a production of a grammatical class of words (this rules out **le mi hasfa**, which LIP does understand, because **lemi** is a word of class LE and that is the only way for the **mi** to make sense in that position) and in all cases except for one limited exception, a space is required only where a pause is possible.

lexicography: Certain large classes of words (the A connectives and relatives, the PA and NI words, the LE words) are very large and do not have adequate formal definitions in the previous sources. I made formal definitions of these word classes, which in many cases do not coincide with the definitions implicit in LIP (the latter are sometimes buggy). In almost all cases,

there is no practical effect on what can be said in Loglan. One bad case is the separate immediately following point. The logically composite PA words as defined in NB3 involve an actual ambiguity (caused by permitted use of initial **noi**) which I repaired.

APA words and kin: Words like *efa* (and more crucially *irau*) present problems of actual ambiguity for the old definition of the language, unless explicit pauses are enforced. Rules requiring such pauses are not clearly stated, though the corpus and some comments suggest that JCB was aware of the difficulty. The complete solution which makes APA words resolvable is rather complex. The APA words themselves might be dispensible, but they are common in the NB3 corpus. The IPA words are not really dispensible in our view. An intermediate solution in which the high level IPA words remain, needed to be closed with pauses or GU, and the APA words are abandoned does seem reasonable to us but is not at this point necessary.

I entirely forbade KA and KI words to absorb following PA words as LIP seems to want to do.

I also had to make provisions to close NI words unambiguously with and without pauses. Details are in the grammar.

acronyms: Acronyms were known to the previous Academy to be a problem. I have been guided in my handling of these by the principle that acronyms are a minor feature of the language and that it is therefore absurd to have to pause between letterals appearing as successive arguments to defend the acronym class. I changed acronyms to be names rather than predicates, which gives them the same boundary defenses as names (starting with name markers or pauses and ending with pauses). Acronyms appearing as dimension in NI words start with a new initial marker word and end with pauses. This leaves one free to use letterals freely for their most important grammatical function, as pronouns, without awkward pauses.

strong quotation: I have an entirely different strong quotation proposal. The original strong quotation is not implementable even in BNF.

serial names: Serial names like **la Djan Braon** were known to be an issue by the previous Academy. The strange solution had been adopted by the previous Academy of having the pause in **la Djan Braon** be of a different length than the other pauses in the language, to allow **la Djan Braon, blanu** to be read as a sentence and **la Djan Braon Blanu** to be read as the name "John Brown the Blue". My view is that serial names are a minor feature of the language which should not have leverage at such a fundamental level.

There is the further general issue of recognizing the beginnings of names. A name word begins with a pause (not necessarily shown as a comma) unless it is immediately preceded by a name marker word (**la** and **hoi** are

examples). This causes names with phonetic copies of name markers in them to be problematic.

We have given a complete solution to both problems. Predicates in serial names must be preceded by **ci**. Name words containing name markers must be preceded in serial names by **ci** (this was already introduced by the previous Academy). Further, names following predicates must be preceded in serial names by **ci**. This allows correct resolution of serial names. All pauses are of the same length, though pauses at the ends of names followed by **ci** or further name words may be written simply as spaces. **La Djan Braon, blanu** is a sentence while **la Djan Braon ci Blanu** is a name.

The solution to the problem of recognizing the start of names is to allow names to begin only in very restricted contexts. The principal hazard in the old language definition was the presence of unmarked vocatives, which meant that a name could start almost anywhere, and so if names could include name markers then large sections of pause-free speech could accidentally be misparsed as names. The elimination of unmarked vocatives has been approved by the Academy. Names thus occur only after name markers (**ci** is a name marker), or after other name words (if they are free of false name markers), or after an explicit pause after a descriptive argument, as in **le blanu, Djan**, if it is name marker free. Complete details are in the grammar.

pauses used as gu:

po sentence forms and the lepo problem:

aci connectives in predicates and logically shared final termsets:

3 Phonology and Orthography

3.1 Introduction

This section is about how Loglan is to be written (both letters and punctuation) and how it is to be spelled and pronounced.

3.2 Alphabet and Capitalization

The alphabet of Loglan has 23 letters **abcdefghijklmnoprstuvyz**, the Latin alphabet without the letters **qwx**. Removing **qwx** is the content of a proposal before the Academy. **aeiouy** are the *vowels*; **aeiou** are the *regular vowels*; the non-vowels are *consonants*; the consonants **mnlr** are *continuants* and may be used in a vocalic manner, in which situation they are referred to as *syllabic consonants*. The names of the consonants are formed by appending **-ai** or **-ei** to the letter (for uppercase or lowercase respectively). The names of the lower case regular vowels are formed by prepending **zi-** to the vowel; the names of the upper case regular vowels are formed by appending **-ma** to the name of

the corresponding lower case form. The renaming of the vowels is a current proposal of mine. The question of what the name of **y** is is open in my mind: **ziy(ma)** is accepted by my parser at the moment, but is irregular in form (**liu ziy(ma)** would not parse!) The language ought to have names for the foreign letters **qwx** but what these names are is open in my mind. (NOTE: names for **qwx** strike me as a definite (small) lack, writing 5/15/2015).

The capitalization convention of Loglan is that any uninterrupted sequences of letters (which may include syllable breaks and stress markers but not spaces or terminal punctuation) may have the initial letter either uppercase or lowercase, and must have all subsequent letters lowercase. The convention of allowing capitalization of the initial letter of the name of an upper case letter when it is embedded in an acronymic word without spacing is attested in existing Loglan text but not currently implemented in my parser. I am inclined to implement this eventually. (NOTE: I have this in mind for imminent handling, 5/15/2015)

3.3 A Note on Styles of Orthography

It was at one time a project to have a phonetic parser for Loglan. We would have thought of this as a separate gadget manipulating something like JCB's phonetic notation in NB3 or Loglan 1. In the event, this is not what happened. In the course of writing my parser, it became clear that a careful analysis of syllables and stresses was implicit in correctly parsing the standard Loglan orthography from the level of letters upward, and adding explicit devices for indicating phonetic features such as pauses (already of course denoted by commas, but not all pauses can be represented by commas under LIP), stress (not explicitly notated at all in the standard orthography) and syllable breaks (represented explicitly by "close-commas" in some contexts in JCB's notation; we have suppressed the close-comma and use hyphens for syllable breaks) seemed natural. We have produced a single parser which admits a continuum of styles ranging from the standard Loglan orthography to a style which I term "phonetic transcript" in which there are no spaces other than those which represent explicit pauses, and these are all marked with commas, and all stresses are shown explicitly (one could further show all syllable breaks explicitly, and for genuine phonetic transcript one should indicate explicitly how optional disyllables are being handled).

3.4 Punctuation

The comma **,** denotes a pause in speech. There are contexts in Loglan orthography where the presence of a mandatory pause can be deduced from the orthography though a comma is not present. It is a design goal in the parser, which I am not sure that I have achieved, to allow an explicit comma to be inserted in any place where a pause is allowed, and certainly whereever a pause is mandatory. A comma is always followed by a space. The use of a close comma to force a syllable break has been abandoned at least for now (if desired, it can be reintroduced as a variant of the syllable break **-** but I advise against it). Wherever a pause is intended, at least a space should be written. The 5/15

upgrade of the parser enforces the restriction that spaces cannot occur in the middle of words, so ideally whereever a space occurs it should be permissible to pause.

The period **.** denotes the termination of an utterance: other terminal punctuation marks **?!:;** are exactly equivalent as far as the parser is concerned. Terminal punctuation ought to be followed by at least two spaces. The parser now enforces at least one space or end of text after a terminal punctuation mark. 5/15/2015: I am considering some other punctuation, such as ellipses, dashes, and (silent!) parentheses and brackets.

The hyphen **-** denotes a syllable break (it may **not** be pronounced as **y** as JCB proposed in Notebook 3). The apostrophe **'** marks a stressed syllable; it may be used instead of **-** (not in addition to the hyphen) to mark the syllable break after a stressed syllable, and it may be used after a final syllable before a non-letter or end of text to indicate that the final syllable is stressed.

The asterisk ***** may be used in two ways: in initial position, it marks utterances which are deprecated or error-infested as Loglan utterances. It may also be used with the same grammar as the hyphen to denote emphatic stress.

Stress markers are always optional. There is no specific notation for a syllable which is not stressed, though it might be useful to add such a marker. (A note for those familiar with the language: of course, when we say that stress markers are optional, we should point out that indications of stress are not optional in some cases: the stress in a predicate word, if not signalled by an explicit stress marker, must be signalled by the end of the predicate word, indicated in this case by a space or terminal punctional mark. In the presence of an explicit stress, the end of the predicate word can of course be determined without additional punctuation.)

I suggest using the stress markers to indicate rhetorical stress in a way the parser can handle. I have been doing this myself.

Spaces between words not occurring after a comma can in most cases be omitted; in certain cases they indicate mandatory pauses which can always be explicitly marked (insert missing comma), or in the case of spaces after predicates may serve to mark penultimate stresses in the preceding word, and can be eliminated if the stress on the penultimate syllable of the preceding word is explicitly marked. Places where spaces are written are often but not always places where pauses are possible in reading the given text, but one certainly should not presume that a space indicates a pause. The aim is that any space where the pause is optional can be omitted and any space where it is mandatory can have a comma added. The 5/15 parser upgrade forbids spaces in the middle of words, and makes it almost true that any place where a space is written is a place where one can pause (for those familiar with the language, there is an exception when an A word is followed by a PA word then a space: the final space cannot be converted to an explicit pause without breaking or changing the parse).

A specific style which should be possible to write for any Loglan utterance is the *phonetic transcript*, in which spaces appear only after commas and all stresses and syllable breaks are explicitly marked (or at least all stresses and

syllable breaks of interest are marked; but if spaces are omitted stresses on predicates become mandatory).

3.5 Pronunciation

Input about actual pronunciation of Loglan sounds from other members of the community is actively solicited.

3.5.1 Regular Vowels

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 *machine*

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).

3.5.2 Grouping of Vowels

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.

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 rule applies to **ii** and **uu** where these are pronounced disyllabically (remark added 5/20).

There are two different rules for grouping long strings of vowels. In a structure word (which will always be a compound attitudinal with an even number of vowels), the vowels are grouped in pairs and each pair is monosyllabic or disyllabic in a way compatible with the rules above. An odd length stream of vowels made of structure words will consist of a one-letter logical conjunction followed by a compound attitudinal.

In a name or predicate word, a long string of vowels without an explicitly given pause or syllable break is grouped using a priority scheme. If the first two letters of the stream make up a mandatory monosyllable, they are grouped together; otherwise, if the second and third letters make up a mandatory monosyllable, the first vowel is pronounced by itself and the second and third are grouped together; otherwise, if the first two letters make up an optional disyllable they may be grouped together or pronounced separately (the parser, absent an explicit syllable break, prefers to group them together), and otherwise the first letter is pronounced by itself; in any of these cases, repeat the process with the remaining stream of vowels until it is exhausted. This is a new proposal of mine superseding rather different rules given in earlier texts.

3.5.3 The Irregular Vowel

The irregular vowel **y** has as its standard pronunciation the "schwa" sound found in English *sof*a. John Cowan suggests that we might sometimes want to give it the value of **oo** in English *loo*k, a suggestion which I rather like. It is very important to note that the English or Russian tendency to convert the vowel in an unstressed syllable to this sound must be resisted.

3.5.4 The Consonant Sounds

The consonants **c** and **j** have pronunciations which are unusual in English.

c is pronounced as **sh** in **sh**oe.

j is pronounced as **s** in *trea*s*ure*.

The consonant **g** always has the hard pronunciation in **g**et.

The consonant **h** usually has its usual English pronunciation but may also be pronounced as **ch** in Scottish *lo***ch** (the former pronunciation of **x**), as for example if it appears at the end of a syllable. (This alternative pronunciation is part of the proposal to eliminate the foreign consonants).

The consonant **n** has its usual pronunciation in English except before **g** or **k**, where it is pronounced as **ng** in *so***ng** (which is also quite usual in English!). The latter sound appears only as a pronunciation of **n** in such contexts. I note that this might also apply to **nh** if **h** has the alternative pronunciation.

The other consonants all have their principal pronunciations as in English.

It is worth noting that **tc** is English **ch** as in **ch**in and **dj** is English **j** as in **j**udge

3.5.5 Syllabic ("Vocalic") Consonants

No consonant ever appears doubled in Loglan, except the continuants **mnlr**. Where the continuants appear doubled, they are used syllabically (in effect, as vowels). We call these doubled consonants *syllabic pairs*. A syllabic pair is never adjacent to another occurrence of the same consonant. We require that a syllabic consonant (a continuant used as a vowel) must always be doubled: the main effect of this rule (which is suggested in Loglan 1) is that the spellings of some names must be changed.

3.5.6 Grouping of Consonants

Consonant clustering is governed by a number of rules.

There is a list of allowed initial pairs of consonants which may begin a syllable. An initial pair will not overlap with a syllabic consonant pair. A Loglan syllable will begin with a vowel or syllabic pair, or with a single consonant (not part of a syllabic pair), or with an initial pair (not overlapping a syllabic pair), or with an initial triple of consonants in which each of the two adjacent pairs of consonants is an initial pair.

The 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**

There is a list of pairs of consonants which may not occur adjacent to one another, even across a syllable boundary. These are called impermissble medial pairs (and the other pairs are called permissible medial pairs).

The 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 impermissble medial triples as well, consisting of **cdz**, **cvl**, **ndj**, **ndz**, **dcm**, **dct**, **dts**, **pdz**, **gts**, **gzb**, **svl**, **jdj**, **jtc**, **jts**, **jvr**, **tvl**, **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.

3.5.7 The Loglan Syllable

It is a curious feature of Loglan as presented heretofore that there is no precise definition of a permissible syllable, and in fact the places at which syllable junctures occur in JCB's phonetic productions are sometimes quite odd. We are much more precise about this, though we believe that we have not thereby in principle much modified the set of allowed words (or in practice modified it at all).

A Loglan syllable consists of three components, only one of which is mandatory. It begins with an optional initial consonant group, continues with a mandatory vowel group, and terminates with a final consonant group.

The initial consonant group is either null, a single consonant, an initial pair or an initial triple in which both adjacent pairs are permitted initial pairs. The

initial consonant group cannot be or overlap a syllabic pair. This treatment of initial triples is stated in Notebook 3.

The vowel group is either a single vowel (**y** can occur), a mandatory or optional monosyllabic pair of regular vowels, or a syllabic pair of continuants. A syllabic pair of continuants cannot be adjacent to another occurrence of the same consonant.

The final consonant group is null or contains one or two consonants, neither of which can stand at the beginning of an impermissible medial pair or triple of consonants (possibly looking ahead into the next syllable). Further, if there are two consonants in the final group, they cannot consist of a non-continuant followed by a continuant (this is a new condition: such a combination would be forced to be pronounced as another syllable). The final consonant group cannot be or overlap a syllabic pair.

A syllable break immediately preceding a vowel must also follow a vowel (any syllable following a consonant group picks up at least one of those consonants).

The position of the juncture in a group of three or more consonants at the end of a syllable may be optional; in no case is it phonemic (there cannot be two distinct words which differ only in the placement of such a juncture). Forcing a syllable break between vowels may change one word to another (this will only happen in proper names).

Of course, further conditions are imposed on syllables depending on the kind of word in which they appear.

It is interesting to note the maximum degree of consonant clustering: CC-CCC is possible in a name or borrowing.

4   Phonetic Classification of Words

There are three main classes of words in Loglan, structure words, names and predicates. This section is concerned with the phonetic shape of these words.

4.1   Structure Words

Structure words (in Loglan, `cmapua`), also sometimes called little words, are mostly grammatical particles. There are some words which are phonetically structure words and semantically predicates, and some which are semantically names (8/1 proposing to have acronymic names rather than predicates).

On the phonetic level, structure words are built out of phonetic units of one of the shapes V, VV, CV, CVV, Cvv-V, where V denotes a regular vowel, C denotes a consonant and vv denotes a monosyllable (optional or mandatory). The units of the shape Cvv-V are currently not used or very little used, but we leave the possibility open. The commonest cmapua consist of a single unit.

If any unit in a structure word is a VV, all units are VV's. These words are the compound attitudinals already mentioned. The compound attitudinals are the only vowel-initial words which do not have to be preceded by a pause.

A unit of the shape V may only appear in initial position, and must be preceded by an explicit (that is, comma-marked) pause. A V by itself falls under this rule. These words are logical connectives of a certain class, and in fact all logical connectives of this class must be preceded by explicit pauses, though some of them have no phonetic reason to be so marked. This will be spelled out in more detail later.

Stress in a structure word is completely free – any, all or none of the syllables may be stressed. If the final syllable of a structure word is stressed and the following word is a predicate, there must be an explicit pause, comma-marked, between the two words. NOTE: The phonetic cmapua which are semantically predicates are supposed to be penultimately stressed, like phonetic predicate words, but this is not currently enforced by my parser (it will be enforced when I get around to it!)

It is worth noting that the actual class of phonetic structure words has hardly any use in the grammar. Grammatically, this class is broken up into individual often very small classes each of which happens to satisfy its constraints. There is only one context in the PEG grammar, following the quotiation article **liu** where the general phonetic class of structure words is used. I believe that in LIP it may never have been used at all, since after **liu** LIP accepts only actual structure words, not phonetically acceptable ones; the latter approach is the one we take.

It is also worth noting that the phonetic units of structure words are not individually necessarily single syllables: a VV or CVV might be a disyllable, and a Cvv-V unit is definitely not a single syllable.

4.2 Names

A name must be made up of valid Loglan syllables and must end with a consonant. The final consonant must be followed by either a comma-marked pause, end of text, a terminal punctuation mark, or a space followed by another name word or the name marker word `ci` (in which case there is a pause in the pronunciation). [It can also be followed by a stress marker followed by any of these things, if it is finally stressed]. I disagree with Steve Rice's opinion in Loglan 3 that the comma after names can be omitted with experience; I believe it is important to reinforce it.

The requirement that names are formed of syllables is a new proposal. The original Loglan rule allowed any string of sounds ending in a consonant. In practice, nothing here has required changes in any Loglan name in use other than enforcing the rule that syllabic consonants must be doubled (which was actually suggested as an alternative by JCB in L1).

An acronymic name is not consonant final but must also have the following pause if not final, which must be explicitly comma-marked unless another name or **ci** follows.

We do not require that names be capitalized, but it is usual to capitalize them.

A name must be preceded either by a pause or by one of a class of name marker words. The name marker may optionally be followed by a pause. The class of name marker words is (currently) **la hoi hue ci gao liu** In orthography, the pauses mentioned here may or may not be comma-marked. Note that it is possible for a name to occur not preceded by a space, but only if it is preceded immediately by a name marker.

A vowel initial name must be preceded by a pause, which must be shown by at least a space and may or may not be comma-marked.

Names are the only consonant final words in Loglan. Thus the end of a name is always readily recognized.

A phonetic copy of a name marker word occurring in a name, such that the part of the name following the phonetic copy is a well-formed name, is called a *false name marker*. A false name marker presents a difficulty for the reader or auditor trying to determine the beginning of a name. The rule is that a name begins as early as possible. To make a false name marker into a true one, follow it with a space or comma marked pause. A further important rule is that a name containing a false name marker cannot occur after another name without being marked (we will see that the marker used is **ci**).

An essay is needed about the pragmatics of avoiding false name marker problems. A possible additional rule is that any name with a false name marker in it must always be preceded by a pause, even when preceded by a name marker. Pragmatically, it is a good idea to pause after a vowel soon after the end of a name, if there is any danger of what follows being misconstrued as a second name. We believe that we have solved the problem of name resolution in our parser, but this did require attention and fine-tuning of a number of grammatical details.

4.3   Predicates

Predicate words fall into two classes, borrowings from other languages and complexes. We describe the class of borrowings first, but we note that when a predicate word is parsed, one first attempts to parse it as a complex, and only after that does one attempt to parse it as a borrowing.

A predicate must resolve into Loglan syllables.

All predicates have penultimate stress, meaning that they are always stressed on the last syllable but one, ignoring syllables not containing regular vowels (only one such syllable may intervene between the stressed syllable and the last syllable), and usually only on that syllable (note for those familiar with the language: it is permitted to stress the final syllable of a nonfinal borrowing djifoa before the **y** hyphen), so of course they have at least two syllables. This helps one to determine where a predicate word ends. All predicates are vowel-final (so they are not names). All predicates contain at least one occurrence of two adjacent consonants (so they are not structure words).

The rules governing the beginning of a predicate word are designed to prevent ambiguity between a predicate word and a structure word followed by a predicate word. If the word begins CC there is no difficulty. A predicate word cannot

contain more than one consonant before the first CC junction, because then the first consonant plus the stream of following vowels could be peeled off as a structure word or words. So the general form of the beginning of a structure word is an optional single consonant, followed by an optional string of vowels, followed by CC, where the CC does not itself begin a valid predicate. The CC cannot be a permissible initial pair broken by a juncture.

A vowel initial predicate appearing in a noninitial position in a sentence must be preceded by a pause (which must be written at least as a space and can be comma marked).

A predicate cannot begin VCCV where the CC is a permissible initial pair.

No predicate can be of the form CCVV.

A predicate cannot begin with a syllable whose vowel segment is a syllabic pair. A syllabic pair cannot follow a vowel in a predicate.

The parser recognizes the end of a predicate either by noticing an explicit stress then counting syllables to the end of the word, or by seeing a space or punctuation ending the word and checking that the syllabification allows the syllable before the previous one (skipping a possible syllable without a regular vowel) to be stressed. This means that in the absence of an explicit stress, some spaces are mandatory which do not represent pauses in speech (though they always occur at points where it is permissible to pause), but rather signal the presence of a stress.

4.3.1 Borrowings

The additional features of a borrowing over and above the general features stated above are that it cannot contain any occurrence of **y** or of any of the disyllables **aa**, **ee**, **oo** which force stress on one of their components (nor of **ii**, **uu** as disyllables). Further, it cannot be a complex, but this is not enforced by the parser rule for borrowings directly, but by attempting to read any predicate as a complex first.

There are additional technical condition on explicit syllable breaks using **-'*** in borrowings whose motivation is described below. The situation CV-CC is not permitted where the final CC is an initial pair, and a CC-C may not have the two consonants separated by the juncture forming a permissible initial pair.

4.3.2 Complexes

A complex is like a structure word in being composed of units which are not themselves syllables and whose interaction with syllabification can be tricky. These units are called *combining forms* officially; traditionally they have been called *affixes*, a deprecated usage. In Loglan they are called **djifoa**.

It is required that a syllable does not overlap with more than one djifoa. This is enforced by the restriction given above on junctures in borrowings: a string formed from a complex only by adding syllable breaks that violate djifoa boundaries will exhibit one of the excluded behaviors, and so will not parse as

a borrowing. The first of the two excluded behaviors for borrowings does occur in complexes.

The djifoa are of the following basic forms (where C represents a consonant and V represents a regular vowel). Each of the djifoa of one of the three letter forms is either an abbreviation for a five letter form (this information is in the dictionary) or is associated with a structure word (also in the dictionary):

CVV: Note that an initial CVV djifoa cannot be followed by another `CV-` form due to the general rules of predicate formation. This is fixed by allowing an optional "hyphen" to be appended to the djifoa. This hyphen may be **r**, or it may be **n** if followed immediately by r initial in the next affix, or it may be **y**. Note that CVV djifoa where the VV is **aa**, **ee**, or **oo** can only occur in final or penultimate position among the djifoa making up a complex, as one of the syllables of such a djifoa must receive the main stress in the word (and if the VV is **ii** or **uu** it can only be pronounced disyllabically in final or penultimate position). The fact that a CVV appearing in final position where the VV is an optional monosyllable can be syllabified in two different ways may cause there to be two possible ways of stressing a complex.

CCV: A CCV never needs to be hyphenated (though my parser doesn't know this).

CVC: Of course a CVC cannot appear in final position among the djifoa in a complex.

A CVC which is initial in a complex will be followed by a **y** hyphen if the next affix begins with a consonant and the CC juncture created would be an initial pair, unless the word is CVCCVV or CVCCCV, which do not need this form of hyphenation (this avoids a CV form falling off the front of the word). The **y** hyphen is a single syllable by itself; in terms of djifoa analysis, it is treated as part of the CVC djifoa. It should be noted that though this is related to the general rule that the part of a predicate beginning with the initial CC cannot be a predicate unless it is the whole predicate, it is not a mere case of this rule, as it is not necessarily the case that dropping the initial CV in such a predicate without the inserted **y** would produce a legal predicate.

This rule is why TLI Loglan no longer has the **slinkui** test. It is not in 1989 Loglan but it is explicitly approved by the academy (with further official modifications in 2013).

A **y** hyphen may also be appended to a CVC djifoa to prevent formation of an illegal medial pair or triple of consonants with the following affix.

CCVCV: In non-final position, the final V is replaced by **y**. If a syllable break is expressed, it is CCV-CV.

CVCCV: In non-final position, the final V is replaced by **y**. If a syllable break is expressed, it may be CV-CCV or CVC-CV – the former is of course allowed only if the CC is an initial pair.

borrowing djifoa: A borrowing djifoa is a complete borrowing plus hyphens; **y** is added before it if it is not initial and after it if it is not final. A borrowing by itself is not a djifoa. These **y** hyphens are regarded as part of the djifoa. A nonfinal borrowing djifoa is optionally stressed on its final syllable before the **y** (which is not the same as the stress on the borrowing itself!); this stress must be expressed if the djifoa is followed by a final monosyllabic djifoa (because in this case the stress is the main penultimate stress on the whole predicate). It is permitted to pause after the subsequent **y** hyphen if the stress is expressed (and to write a space or pause there). Recall that **y** never appears in a borrowing; this makes it clear that borrowing djifoa can be resolved.

A complex is a word which satisfies the general conditions to be a predicate and resolves into djifoa (where any hyphens used are regarded as part of the affix). A single five letter djifoa is a complex (a primitive predicate).

4.3.3   The zao construction

John Cowan has proposed this as an alternative to the use of borrowing affixes, and I quite like it as an option. A sequence of predicate words separated by the word **zao** is grammatically a predicate word. There is no grouping in this construction any more than there is in the basic predicate construction. It is permissible for one or more of the initial items in a sequence of words linked with **zao** to form a predicate to be CV or CVV affixes.

I see one use of this construction as the ability to paraphrase a complex whose structure may be unclear to someone.

NOTE: I have a preliminary suggestion that all CVh affixes can be regarded as associated with the CV structure word with which they begin.

5   Word Forms

This section deals with details of Loglan that are for the most part not manifest in the previous official formal grammar. The word class definitions are nowhere actually given formally; they are implicit in tables internal to the old interactive parser which are not human-readable and clearly have bugs.

Our program in desigiing the PEG parser wass to parse Loglan from the level of letters upward, and as a result we have had to mandate exact formal definitions for these word classes, which in some cases are clearly not exactly the same as those implicit in LIP. Details will be seen below.

Quotation constructions and other constructions which import foreign text are handled in this section. My implementation of strong quotation is a completely new proposal.

3/28/2015 Cyril points out that it is very important to articulate the concept of "word" formally. As JCB says in NB3, the defining characteristic of a word is that one cannot pause in the middle of it. He says this in the abstract, but then does not give us any formal definition of cmapua words: the definitions of word classes in LIP are part of the internals not expressed in the formal grammar we inherited. We are told that in Lojban there are no cmapua words with more than one syllable in this sense. This is not the case in TLI Loglan. Certain cmapua classes are genuinely classes of words, in that one cannot explicitly pause in the middle of a production of this class. The 5/15 parser upgrade forbids words to be written with spaces in them (so one cannot write **le na hasfa**: **lena** is a word). Earlier versions of my parser and LIP permit **le na hasfa** but note that they did not permit **le, na hasfa**.

It may very well be that with further work we could achieve the situation reported in Lojban where a stream of one-syllable cmapua is understood without reference to any pauses that may occur between syllables, so that there are no multi-syllable cmapua which are words in this sense. We further note that we do regard it as unfortunate when the placement of a pause in a stream of cmapua syllables materially affects meaning, though we observe some situations where this seems difficult to avoid (the classic `le, po` problem has now been resolved, though). We do not however regard the concept of "multi-syllable cmapua word" as alien to Loglan: JCB clearly envisaged there being such words.

5.1 Pauses

Some pauses in Loglan are syntactically significant. A pause cannot be significant if it is required for phonetic reasons. The pause after a name word is not significant. The pause before a vowel-initial word or an A series logical connective is not significant. A pause after a stressed syllable before a predicate is not significant. We will use the phrase "significant pause" to refer to pauses which can be syntactically significant. Further, a pause followed by **y** then a second pause or by `cuu` is not significant (these two constructions allow one to cancel a pause which might otherwise have grammatical effects – and of course these can be used to pause for effect!) The pause cancelling construction is a proposal.

There is a general rule that a pause, expressed or unexpressed, cannot occur in the middle of a word, and that spaces are not written in the middles of words.

5.2 Structure words

We begin by considering the many classes of structure words.

5.2.1 Logical connectives for sentence components

There are numerous parallel classes of logical 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.

The basic series of connective roots is **a, e, o, u, 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 connectives. The prefex **nu** may appear initially to a logical connective word of the basic series; it may only appear if followed by **u** or **no**. The root taken from **a, e, o, u, ha** (possibly with prefixed **no** and/or affixed **noi**, i.e., an A core) follows this. Finally a complete PA word (a tense in the broadest sense) may follow as a suffix; if a PA suffix is present it must be followed by a significant pause or by gu; an optional significant pause or **gu** may appear in the absence of a PA word. An A word may not be followed by whitespace then a PA word then an explicit pause. If an A word is followed by a PA word then a space, this final space cannot represent a pause (this is the only exception to the general rule that where a space is written, a pause is permitted).

All A words are preceded by explicit comma-marked pauses (which cannot be significant 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 because the related IPA words, which present much the same difficulties of termination, are clearly not dispensable without doing some violence to the corpus.

The ability to attach a PA word as a suffix is under review; it is clear that the following pause or **gu** is necessary to avoid ambiguity. The following pause or **gu** is a proposal; it seems that JCB was intermittently aware that such pauses were necessary to keep an IPA word from being confused with an A word followed by a PA clause. The pause or **gu** without accompanying PA word is an option which may be useful to disambiguate linked modifiers.

NOTE: it remains possible in my mind to preserve the IPA, ICAPA words and eliminate the APA words. But solutions seem to be in place for the difficulties.

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 **nu** converts a logical connective to its converse. 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 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 **gu** 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. which is not a significant pause, just as in the case of A connectives.

The CA connectives are another related class. They are not preceded by pauses. The CA root forms are **ca, ce, co, cu, 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, 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; we will see if they are useful.

5.2.2 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, followed by **gu** or a significant pause. All words of this class are preceded by a phonetically mandated non-significant pause.

A word of the class ICA consists of I followed by a CA connective word. This is a logical connective acting between sentences. Because it is vowel-initial, it must be preceded by a non-significant pause.

An I or ICA word cannot be followed by whitespace then a PA word (an explicit pause is needed to separate an utterance 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** (noting that a final PA word will not be closed with **gu** or a pause in this case).

The fact that a connective like **irau** or **icafa** which ends with a PA word must be followed by a significant pause or terminated with **gu** is a significant point of the grammar. The closure with **gu** is a new proposal here; the need to at least pause after these connectives seems to have been noticed by JCB in some contexts.

5.2.3 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** (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!

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

5.2.4 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),

zoa: (double prime),

zoi: (prime),

 pi: (decimal point),

re: (more than half of (quantifier)),

ru: (enough of (quantifier)),

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),

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).

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 of numeral units consisting of a numeral (any word of class NI0 but this really makes sense only for the digits) followed either by `ma` or by `moa` and an optional further digit. D `ma` means D followed by two zeroes; D `moa` means D followed by three zeroes. D `moa` n

23

means D followed by 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.

A numeral block is a sequence of atomic quantifier words, numeral units and RA words in which a RA word if final is the only element.

A general quantifier word has a quite complex definition. It may optionally begin with **ie**. It is then followed by a numeral block, which may optionally be linked with more numeral blocks by CA cores (a CA root with optional negative prefix and/or suffix) [this structure of CA-linked numeral blocks is called a NI core]. This may optionally be followed by an acronym which must start with the marker **mue**; 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**. A pause may be necessary to distinguish a NI word from a following NI word; to avoid this one may terminate a NI word with **pui** or a pause followed by **pui** (**pui** is always followed by another NI word.)

The word **ie** used as a prefix here is an interrogative word meaning which. 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.

The device **pui** for making a break between adjacent NI words without a pause is a proposal.

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 avoid. The word "mex" (abbreviating mathematical expression) is used in the grammar section for quantifier words.

5.2.5 Letters, acronyms, and pronouns

A Loglan upper case consonant letter is C**ai**. A Loglan lower case consonant letter is C**ei**. A third series C**eo** is provided for lower case Greek letters.

A Loglan lower case vowel has the form **zi**V, and the upper case form is **zi**V**ma**. The old style forms V**fi** and V**ma** are currently supported in the parser but deprecated, and (NOTE) it is proposed that they be eliminated.

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: (this [of situations]),

tio: (that [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 nth description back in the text not already bound to a pronoun will be bound to the nth letter on this stack when needed. It seems rather baroque but very simple cases can surely be used correctly.

The general class of pronoun words consists of letters or other pronouns, optionally suffixed with **ci** followed by a quantifier word (usually a small numeral!) It is very important to notice that for us a pronoun is a **single letter**, possibly suffixed with a numeral. 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.

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 is grammatically far more important than any use of sequences of letters as pronouns or acronyms.

Further letter words, which may be used as pronouns, but to which we may not attach numerical suffixes (? I may want to allow this), are generated by **gao** followed by a single well-formed word, either a name, a predicate, 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.

An acronym is a sequence of letter names (possibly abbreviated in the case of vowels to zV – not to just V – which eliminates distinctions of case of course), 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 noted above) 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. A pause, terminal punctuation, or end of text is required after an acronym (so it can never attempt to consume a following letteral pronoun).

5.2.6   Tense/location/relation operators

The root words of this class (which we call PA words 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),

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

fea : (???I cannot find any indication of what this means, though there is a reference to it in text of trial.85),

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),

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: (???),

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),

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),

rui : (???),

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)

and the 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)** which can be prefixed with **nu**, **no**, or **nuno** to give additional forms which we call KOU cores (a root is also a core).

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.

A compound PA word begins with an optional numeral or quantifier, followed by a string of PA roots or KOU cores taken from the lists above (recalling that KOU cores may include certain prefixes), optionally linked to further strings of PA roots/KOU cores by CA cores, then may optionally be closed with one of the qualifiers **za, zi, zu** (to see the effects of these qualifiers on tense and location operators, see the dictionary) or with **juu** or a pause followed by **juu** (the **juu** forms must be followed by another PA word; this is a device to avoid necessarily having to pause when one complex PA word is followed by another).

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 words can be used as prepositions (followed by an argument) or as tenses in the broadest sense (followed by a predicate). 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, it is generally illegal for it to be replaced by a space followed by a PA word : where a PA suffix is legal, it cannot be replaced by a following PA word (the rule for the optional PA suffix on an A word is different, and stated above; an A word without a PA suffix can be followed by a separate PA word as long as this is not in turn followed by GU or an explicit comma pause). It is illegal for a PA word without one of the optional closing components to be followed by a space then another PA word. These are all rules designed to defend the condition that spaces are forbidden to occur in the interiors of words.

The only spaces in the orthography which cannot be pronounced as explicit pauses are spaces following a PA word which follows an A word.

5.2.7 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.

vuza: 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.

vuzu: 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 word (this is true in my grammar, and experiment confirms that this happens with LIP as well). In a word such as **fanazu**, what does the **zu** qualify? It seems most reasonable to suppose that in a word **pacenazu**, the **zu** qualifies both conjuncts. The current grammar does not allow logical conjunction of PA cores with different **zV** qualifiers to form words.

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

5.2.8   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 objects(s) understood from context with property X...

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

la: The article for proper names.

These are the root words. Complex words of this class can be formed by following the root with an optional pronoun followed by an optional PA suffix. Spaces may not intervene between these components.

The name constructor **la** and complex articles derived from it appear in the list above but appear 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 some quantifiers (the NI cores) 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. These are forms to which I admit having paid very little attention.

5.2.9   Constructions involving alien text and related articles

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).

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 of the various classes in this section. 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 pause may sometimes be required to terminate a quoted word where you want it terminated. **niu** may be used instead of **liu** to explicitly signal that a quoted word, though phonetically acceptable, is not a Loglan word. I have just installed the correction allowing **liu ziy** and **liu ziyma** to parse.

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. Whereever 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 termnology 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 **cii** before each subsequent block. The level of quotation (quotation, quotation of a quotation, etc) can be modified by suffixing a numeral directly to **lie** or **cii** (and it can change from block to block; where a change occurs the numeral must be explicitly given (the only case where **ne** is required)). This device was added because this technique of quotation would become really burdensome if nested quotations were made explicitly. This is an entirely new proposal. The original strong quotation method is not PEG parsable (it is not even BNF parsable) and I think has other weaknesses.

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

We support in the grammar without necessarily approving (also without necessarily disapproving; I know some Keugru members do not like them) the qualifiers **za** (text) and **zi** (sound) for quoted forms. The qualifier is to be affixed to the initial quotation article (after any suffixed numeral in the case of **lie**).

5.2.10   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**

The particles **je** and **jue** mark tightly bound arguments (or modifiers, according to a proposal).

The JI words

jie: (restrictive set membership),

jae: (nonrestrictive set membership),

pe: (general possessive),

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

ja: (which/that (is) nonidentifying

construct subordinate clauses from arguments, modifiers or predicates.

The JIO words **jio, jao** construct subordinate clauses from sentences (resp. identifying, nonidentifying)

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),

jui: (lessers),

 neu: (conditions/circumstances/fields),

pou: (products/purposes),

goa: (greaters),

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, but they have the same grammar as case tags. 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.

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, fuo, 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, puu, zoo** and the long-scope **po, pu, zo**. In each set, the words form predicates for events, properties, and quantities respectively.

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

5.2.11 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.

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.

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 :-)

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 : 9maybe (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 preparser 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.

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**. [I have it in mind to make these words be additional name markers, so that one can say **Loa Djan** as well as **Loa, hoi Djan**; this would also require new freemod forms]

The attitudinal, discourse and social words (class UI) can be negated by preceding them with **no** or following them wiht **noi**.

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

5.2.12 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.

5.3 The Large Word Classes

There remain the large classes of predicate and name words.

5.3.1 Predicate words

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 predicates semantically, though they are structure words phonetically. They form a grammatical class BI of identity predicates.

NOTE: I am thinking of adding to this class all the forms obtained by prefixing **nu**, giving converse operators.

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),

bua: (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! (some essay will be needed here).

The class PREDA of predicate words includes the last list.

The class PREDA includes quantity words suffixed with **ra, ri, ro** to form numerical predicates (cardinal, ordinal, quality ordinal, respectively). A predicate **tora** is a two place predicate, X is a two element subset of Y; **tora** is a two place predicate, X is the second term in series Y. I do not know what the quality ordinal predicates are supposed to be like (homework for me to do). The numerical predicates should be penultimately stressed just as ordinary predicates are [the grammar currently does not enforce this].

And of course the class PREDA includes the predicate words in the phonetic sense of the first section.

5.3.2 Borrowing predicates

A section which is owed.

5.3.3 Making complex predicates

A section which is owed.

5.3.4 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 statement 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.

6 Grammatical Constructions

This part of the document is fairly closely based on the last official Loglan BNF grammar which underlies LIP. There are changes, major and minor, which I will mention as we encounter them. I started trying to write it in the order presented in the grammar, and this is simply wrong. I have taken an alternative approach

working through trial.85 backwards, more or less, hoping that this will give a more top down view.

6.1 Sentences and Utterances

This corresponds to the last part of the trial.85 or loglanagenda document, which discusses sentences and utterances.

6.1.1 The most basic sentences

The most typical Loglan sentence consists of terms (a list of arguments and/or modifiers) followed by a predicate: it is important to note that the predicate may include a final list of arguments, so this is the form of an SVO sentence. The set of terms is even more usually a single argument (the subject of the sentence) but more than one argument may be followed by the predicate. The initial list of terms must include at least one argument, or the sentence will be understood as an imperative. **Da de blanu** has the same meaning as **da blanu de**, "X is bluer than Y".

Another alternative is the gasent, a VOS construction, which consists of an optional initial **no** of negation (which should not be followed by a pause if intended to be part of the gasent, though it may be followed by other free modifiers) followed by **ga** or a tense marker, followed by a bare predicate without tense marker, followed by **ga** followed by more terms. The bare predicate may include final arguments: the arguments after the **ga** are initial. **Na blanu de ga da** means the same as **Da na blanu de**, "X is now bluer than Y".

It is possible for the final component **ga** + terms to be omitted, giving a sentence like **Ga blanu** or **Na blanu**: (It is) blue. "It's raining" can be said in this way: **Na crina!** where just **Crina!** would be an imperative, meaning something like "be a raindrop!". When the final **ga**+ terms is omitted, a missing **ga ba** is understood.

As explained below (class sen1), an unmarked predicate followed by terms and possibly preceded by one or more terms not including any arguments (9/12/2015) is an imperative sentence: **Donsu ta mi**; "give that to me"; **Na la Ven, donsu ta mi**: "At nine, give that to me". Marking the predicate makes it a declarative sentence with an indefinite subject: **Fazi donsu ta mi**, "Someone will shortly give that to me". This is to be understood as **Fazi donsu ta mi ga ba**. The sentence **Fazi donsu ta mi ga la Djan**: "John will shortly give this to me".

It is also possible for one or more modifiers to appear before a gasent.

NOTE TO SELF: the freemod after the initial no in gasent might profitably not be allowed to be a pause, to avoid notional ambiguity with neghead. This might apply in other places. I'm not sure there is any semantic ambiguity.

All three of these forms are options in the grammar rule "statement": it looks for a gasent first, then a gasent with initial modifiers, and only then for a sensible SVO sentence. Notice the reordering of what is going on by the hearer

when **di fa donsu de...** (which sounds as if Z will give Y to ...) is completed **di fa donsu de ga da**.

NOTE TO SELF: in NB3 the [terms gasent] discussion envisions the terms as modifiers and does not discuss the order of arguments if the initial terms are arguments. I think the mental readjustment is less radical if the terminal **ga** just tells us that the argument that follows it is the first argument, rather than reversing the order of the other two blocks of terms.

Underlying this is the form $Px_1, x_2, \ldots, x_n$ of a sentence in logic, with a predicate (verb) followed by a list of objects. To accommodate the most typical word order in natural languages, this was changed to $x_1Px_2, \ldots, x_n$. In what I regard as a much more dubious decision, this was grouped $x_1[Px_2, \ldots, x_n]$, with following arguments incorporated into the predicate. Further modifications are that the x_1 may be replaced by a series of arguments and that the list of arguments may be padded with modifiers (tense/location/relative clauses) which may appear in any position, before, between or after the arguments.

The maneuvers to move an initial argument or segment of arguments to the end are part of a scheme for achieving all the possible orders of subject verb and object(s). This scheme is not completed yet: there is a further device for fronting a final sequence of arguments (allowing arguments which are final to a predicate to appear first) which does not appear quite yet because it distributes over logically connected sentences, as the subject or initial arguments moved to the end by **ga** do not.

6.1.2 Logically connected basic sentences (and final arguments moved to the front)

The next group of sentence forms to be introduced are logically connected forms. A forethought connected form is a keksent – I will defer describing this for a moment.

A sen1 is a statement, solitary predicate without a tense marker (an imperative sentence, as noted above), or keksent.

A sentence is a sen1 followed by zero or more logical connectives of class ICA followed by a sen1: **Da blanu, ica de blanu, ice kukra!** "X is blue, or Y is blue, and run!" We must note semantically that these group to the left: **(Da blanu, ica de blanu), ice kukra**. This is important when different logical connectives are used together.

An uttAx is a sequence of terms followed by **gi** followed by a sentence, possibly closed with a **gu** or significant pause. The terms are final to the predicates involved and distribute over all the logically connected subsentences if indeed the sentence is logically connected (which is why an ability to close the sequence is needed). Head terms can also be connected with **goi**: these are quantifier prefixes and require separate extended semantic discussion!

De gi da blanu Simple OSV order without logical connection issues, "X is bluer than Y".

Di gi mi cluva, e tu donsu de means "I love Z and you gave Y to Z". If I want to follow this with an afterthought connective and a sentence without Z

as a final argument, I need the pause.

Ra ba goi, ba cluva mi "Everyone loves me" is an example of quantifier prefixes. Of course one can say **Raba cluva mi** in this very simple case.

NOTE: A semantic point (and a proposal): our Sources dictate that the last argument before **gi** must be the actual last argument of the predicate, so that we can skip middle arguments. I regard this as a bad idea; one reason for this is that there are predicates which have many arguments, the last of which may be very obscure to the speaker and/or listener. Instead, I propose that the default position for the final arguments be that the first one in the block immediately follows the last argument appearing in the sentences following (as in my example); if the first argument in the **gi** block (after any arguments with semantic case tags) is marked with a positional case tag, it will take that position in all following sentences and following arguments will be in the positions following the explicitly marked one. This will allow the desired argument-skipping effect.

A keksent is an optional negative **no**, followed by a word like **ke**, followed by a sentence or uttAx, followed by **ki** followed, surprisingly, by the very general uttA1 class of utterances described below, which does include the various sorts of sentences given so far. **Ke mi vizka tu ki mi cluva tu** "I see you and I love you".

NOTE TO SELF: the actual way that keksent is written in the grammar seems to have strange options in it, and I dont know why uttAx is inlined rather than being mentioned.

6.1.3   Free Modifiers and Utterances

I will now move up a level to general forms of utterances.

The first topic is *free modifiers* (freemods). These are a rather miscellaneous collection of constructions which have the feature that they can be inserted into a Loglan utterance almost anywhere. In almost all locations in between elements of a Loglan rule, a free modifier may appear. A position before one of the closing forms (**gu** and the special terminating forms **gue, gui, guo, guu, geu** is not regarded as a medial position where a freemod can be expected to be allowed. A closing form may itself always be followed by a free modifier (modifying the construction which it closes as a whole). They may not usually appear at the end of a grammatical construction, though in some cases free modifiers which are not pauses may so appear. The point is that a comma pause is one of the manifestations of a free modifier, and so free modifiers in their general form cannot appear in contexts where a pause may be significant, which includes the final position in many if not most grammatical constructions. A free modifier is generally attached to what it follows, so free modifiers almost never appear at the beginning of grammatical rules.

The varieties of free modifier follow:

negative attitudinals: phrases like **no ui** and **no sia** fit in here. It is important to notice that this *no* has no logical negative effect. **No ui mi hijra**,

"Unhappily, I am here". A pause here breaks this effect. **No, ui mi hijra** seems to mean "It is not the case that I am happily here".

attitudinals: Words like **ui**, or importantly the word *ei* that turns a sentence into a yes/no question.

"smilies" **soi** followed by a predicate of the descpred class, suggests an action or attribute of the speaker. **soi crano** is a quite literal translation of `:-)`

register markers: indications of attitude toward the one addressed such as **die**, dear.

negative register markers: indications of negative attitude toward the one addressed, such as **no die**. NOTE TO SELF: should NODIE be a separate class which NO1 is guarded against?

parenthesized utterance: **kie** followed by any complete Loglan utterance followed by **kiu**. A side remark.

inverse vocative: **hue** followed by a name, a statement, or an argument list, indicating who is speaking.

vocative: This is a separate grammatical form introduced later. **Hoi Djan** is an example. A reform of the language forbids the simple use of a name by itself as a vocative, even when preceded by a space. For reasons, see discussion of the "false name marker problem".

cancelpause: A sequence of the form [**comma**] **cuu** or [**comma**] **y** [**comma**] (the second form is borrowed from **la Sorme Lengu**). The intention is to allow an unintentional pause which might otherwise be construed as significant (equivalent to **gu**) to be cancelled harmlessly. It could also be used to support pausing for effect.

pause A comma pause, not significant.

scare quote: an optional numeral followed by **jo** (**jo** being equivalent to **nejo**) signals that the preceding word(s) (the number of words suggested by the numeral) are not to be taken literally. **Ai tu fremi jo mi Certainly you are my "friend"**. How to ooze insincerity. An explicit pause will allow a numeral to be put in scare quotes. **Ti ne, jo zavlo** "This is "one" bad thing".

Now we commence the treatment of utterance forms.

An uttA class utterance is a connective, the word **ie** (how many?), or a number. These can only occur as utterances as answers to questions.

An uttA1, a very general class of utterances already mentioned in the keksent class above, may be a sen1 class sentence, an uttAx class sentence **no** by itself, a tightly bound argument list beginning with **je** or **jue**, an argument modifier (subordinate clause, class argmod), a list of terms followed by a keksent (NOTE

TO SELF: this I do not understand; it is there in trial.85), a list of arguments, or an uttA. This may be terminated with a period or other final punctuation.

My best guess is that in the terms followed by the keksent, the initial arguments or modifiers distribute over all logically/causally connected components of the keksent. It might be intended that the initial terms be a modifier, but the rule is there with general terms.

I think the uttA1 class, which includes a lot of utterance fragments, often serves to provide a form for answers to questions. But it does include the sen1 and uttAx classes of complete sentences so it can include quite general utterances. And it enters into the makeup of the keksent class. Some of the forms it permits for keksents are very weird. **Ibuo nukouki mi no nu fatru ki lo aurmo** "But I don't care, because gold!", a translation of part of a gaming joke my son likes.

There are now various layers of utterance up to the full Loglan `utterance` class.

An uttC is an uttA1 preceded by one or more **no**'s set off from the following utterance by **gu** or a pause; an initial **no** will otherwise be absorbed into some shorter structure at the beginning of the utterance.

An uttD is one or more uttC's linked by the afterthought ICI connectives.

An uttE is one or more uttD's linked by the usual ICA connectives. NOTE TO SELF: there is a sort of ambiguity here. The class sentence is formed by linking sen1 sentences with ICA connectives, as in **Mi blanu, ica tu redro**. But this utterance standing by itself will be read as an uttD. Food for thought, though there is no true ambiguity I suppose: both utterances mean the same thing.

An uttF is one of more uttE's linked by I class connectives. Notice that causal connectives like **ikou** are of the I class not the ICA class and so will bind less tightly than the ICA logical connectives.

The Loglan utterance can be one of the following, with the further restriction that an utterance cannot begin with the little word **ge**:

free modifier initial utterance: A non-pause free modifier followed by another utterance. This is basically the only free-modifier-initial construction in the language.

free modifier alone: A free modifier alone, with the same restrictions. It checks for the following utterance first. A period or other terminal punctuation will close this.

ige construction: An uttE followed by an IGE afterthought connective, which links the first utterance to the entirety of what follows (no left grouping). This rule is the reason that an utterance cannot begin with **ge**, to avert ambiguity.

i (or another I word) followed by a free modifier: just what it says. A period will close it.

an uttF: Just what it says.

i (or another I word) followed by an uttF: Just what it says.

6.2 Predicates

NOTE TO SELF: I do propose to allow **nu** to appear before the BI class predicates. Converses of these predicates make perfectly good sense. They are all binary, so the rule should allow exactly **nu**. The simplest way to do this may be to allow initial **nu** as an optional component of the words of this class, making this a lexicography issue rather than a grammatical issue.

At least initially, we will discuss construction of predicates from the bottom up.

6.2.1 The basic building blocks of predicates: predunit classes

The class predunit1 consists of predicates which are in a certain sense atomic (basic building blocks).

1. Predicates of the form **sue eep** (onomatopoeia) or **sao antidisestablishmentarianism** (foreign predicates). Details of these forms are discussed in the lexicography section. They are semantically quite different from each other but share the trait of being formed using a little word followed by alien text.

2. a conversion (or reflexive) operator followed by **ge** followed by a despredE (a flavor of descriptive predicate described below) closed off optionally by **geu** (or the archaic **cue**). This is a sort of parenthesis operation (with conversion) allowing a more complex predicate to be treated as a basic predicate building block.

3. a conversion (or reflexive) operator (such as **nu**) followed by a predicate word (this is just a basic predicate with arguments reordered).

4. The parenthesis form without a conversion operator: **ge** followed by a despredE closed off optionally by **geu** or **cue**.

5. An abstraction forming word like **po** followed by an uttAx sentence closed off optionally by GUO (**guo** or **gu** or a significant pause).

6. An abstraction forming word like **po** followed by a "sentence" closed off optionally by GUO (**guo** or **gu** or a significant pause). This form and the previous one are part of a repair to the language which I made recently: in trial.85 the uses of these kinds of predicates are incredibly (and unnecessarily) constrained.

7. The predicativizing little word **me** followed by an argument closed off optionally by **gu** (optionally flanked by explicit pauses on one or both sides) or a significant pause [i.e., a gap].

8. a predicate word (see above in the lexicography section).

Any of the above forms of predunit1 may include free modifiers in all medial positions (other than the first) and any of them may include a free modifier which is not a pause in final position.

A predunit2 is formed by affixing one or more **no**'s (possibly followed by free modifiers) to the front of a predunit1.

A predunit3 is a predunit2 followed oprtionally by a list of arguments of class linkargs (tightly bound with **je** or **jue** as we will describe).

A predunit is either a predunit3 or a predunit3 preceded by a short-scope PO operator such as **poi**. The predunit is an important level to pause at, as this is exactly the sort of predicate which can appear as a unit in a serial name like **la Djan ci Blanu** ("John the Blue"). More to the point, we can have **la Djan ci Blanu Je Tu**: John the Bluer-than-You, or **la Djan ci ge Cmalo Hasfa** (John the Small House), but not *¹la Djan ci Blanu Tu** or *¹la Djan ci Cmalo Hasfa**. It is important to notice that modification of one predicate by another can occur in a predunit only inside a **ge...(geu)** block.

A kekpredunit is a (possibly multiply negated) forethought connected pair of predicates (in the most general sense to be seen at the end of this section). The form is one or more **no**'s of negation followed by a word of class KA followed by a general predicate followed by a word of class KI followed by a general predicate. At every juncture except after the general predicates a free modifier may be inserted. A simple example: **no ke blanu ki cmalo**, "not both blue and small".

6.2.2 Description predicates

These are predicates intended to appear in descriptions (as components of "noun phrases") rather than those which appear as "verbs". We will see below the contexts in which they are used.

A despredA is the most tightly bound metaphor construction: it is a sequence of predunits and kekpredunits separated by **ci**. **cmalo ci hasfa**, "small house". **cmalo ci nirda ci hasfa**, "(small bird) house". All metaphor constructions group to the left. To say "small birdhouse", **cmalo ci ge nirda ci hasfa**. Note of course that all these phrases make perfect sense with all instances of **ci** omitted: these phrases would only normally be used embedded in a more complex construction. Free modifiers can appear before and after **ci**. [There is a check preventing the **ci** here from being followed by a name word, as **ci** has an independent use in serial names; I am not sure what this defends against, but I am sure I had something in mind].

A despredB is either a despredA or the little word **cui** followed by a despredC followed by a CA word followed by a despredB. An example is **cui cmalo bekti ca groda** "a small thing or a big (thing)". Free modifiers are allowed next to the CUI and CA in medial positions.

A despredC is a chain of despredB's. This is a special version of the basic metaphor construction, as used between a CUI and a CA in the previous rule. No free modifiers (and no pauses) are allowed between the despredB's.

Free modifiers **could** be allowed, but this might cause confusion with the more general case of despredE below where this cannot be allowed.

A despredD is a chain of despredB's linked by CA words (free modifiers allowed before and after the CA words). This is top level logical connection with CA words. These are grouped to the left.

A despredE is a chain of despredD's with nothing intervening (no inserted free modifiers or pauses allowed; a pause would break off a final sentence predicate); this is the top level metaphor construction, grouped to the left. **cmalo nirda hasfa** is a house for small birds and **cmalo ge nirda hasfa** is a small birdhouse. The despredE class is of special note as being the sort of predicate which can be enclosed in **ge...(geu)** to form a predunit1.

le cmalo nirda, hasfa means "the small bird is a house"; **le cmalo, nirda hasfa** means, "the small thing is a birdhouse". These are both deprecated sentence forms: one really wants to use **ga** to mark the verbs here. But we do have reasons to defend the formal possibility of making these distinctions.

A descpred, the top level class of predicates used in descriptions, is either a despredE or a despredE followed by **go** followed by a descpred, where the order of modification is reversed: **nirda hasfa go cmalo** is a small birdhouse. Free modifiers including pauses are allowed before or after the **go**.

Detailed examples of metaphor constructions using all the indicated features are owed (and can be found in L1 and NB3).

6.2.3   Sentence predicates, first pass

The basic sentence predicates belong to classes senpred1-4 exactly analogous to despredA,B,D,E, with the following differences:

1. a senpred1 contains no kekpredunit components (query: why do we not allow non-initial kekpredunit components?)

2. a senpred2 is a senpred1 or a despredB beginning with **cui**.

3. a senpred3 is either a senpred2 or a senpred2 linked by a CA word to a despredD.

4. a senpred4 is a senpred3 or a senpred3 followed by a chain of despredD's. Free modifiers or pauses may appear between components of this chain.

The motivation of these rules is that we do not want kekpredunit-initial chains of modifiers and we do allow pauses between items in a chain of modifiers. The use of description predicates as components does rule out some pauses that might be harmless.

A sentpred is a senpred4 or a senpred4 followed by **go** followed by a barepred. This is more general than a descpred because a barepred may have a list of arguments attached (loosely rather than with JE/JUE).

6.2.4   Sentence predicates, second pass

Here we introduce a black box: a termset is a (quite complex as we will see) argument list which can be attached to a predicate. In **cluva la Djan**, "love John", **la Djan** is a termset. These can be much more complicated, but their internal details do not enter into the grammar of predicates (though they complicate the semantics!) A termset is an argument, or list of arguments, or a structure built by logical connection of simpler termsets.

A kekpred is a kekpredunit followed by a chain of despredD's with optional intervening free modifiers. (I am a little puzzled by this: it appears that as of trial.85 we are allowed to have kekked head modifiers in sentence predicates, with precautions to prevent their falling apart: notice that a kekpred is a description predicate in form. I have now looked at trial.85: note 75 characterizes the desired behavior, which my parser does indeed replicate; my objection is that it appears that the trial.85 grammar is demonstrably ambiguous as written once this change is made; I can exhibit distinct parses of the same utterance; this is not a problem for my grammar or indeed for my understanding of the intention here, as the PEG as written appears to do exactly the intended thing as described in note 75; compare the parses of **da ka redro ki blanu hasfa** and **da ka redro ki blanu gu hasfa**).

NOTE: I do wonder if this change allowing kekked head modifiers means that the sentpred/descpred distinction should be seriously rethought. It is still important not to allow pauses in metaphor strings in descriptions, but that reluctance could also be eliminated by not allowing certain already deprecated sentence forms. This note is not saying that the grammar doesn't work, but that it might be seriously simplified.

A barepred is a sentpred or kekpred followed by an optional termset (which may have a free modifier before it). This is the class which can appear after **go** in a sentpred.

A markpred is a PA word or **ga** followed by a barepred. This is a predicate with a tense marker, with the option of the null tense marker **ga**.

A backpred1 consists of one or more **no**'s of negation (with following optional free modifiers) followed by a barepred or markpred. The caveat applies to each **no** of negation that it does not start a predunit2: in **no blanu hasfa** the initial **no** is captured in the predunit2 component **no blanu**. **no blanu hasfa** means "is a non-blue house". **no ga blanu hasfa**, where the negative is not captured, means "is not a blue house".

A backpred is either a backpred1 or a structure built by linking backpred1's with ACI afterthought logical connectives and optionally adding a termset (shared by all the logically linked backpred1's), or a structure built by linking general backpreds with ACI connectives and optionally adding a termset (shared by all the logically linked backpreds).

A predicate2 is either a backpred or a structure built by linking backpreds with A afterthought logical connectives and adding a termset (shared by all the logically linked backpreds) or a structure build by linking general predicate2's with A afterthought connectives and adding a termset (shared by all the logically

linked predicate2's). A predicate2 cannot begin with **ge**, nor can a backpred following an A connective begin with **ge** (to defend the AGE connectives); nor is there any reason that it should.

Both ACI and A logical connectives group to the left.

This approach has the same practical effect as the trial.85 approach in most cases, but is quite different in detail (and in background theory). First of all, the ACI connectives are fully privileged logical connectives binding more tightly than the A connectives. Secondly (and perhaps most strikingly) no distinctions are drawn between marked and unmarked classes; these distinctions seem to be unnecessary even in trial.85. Thirdly, the handling of logically shared final termsets is rather different. The trial.85 solution is quite lovely, but extremely hard to implement in a PEG. It seems most unlikely that a layering of logically shared final segments of termsets which could not be handled by the rule we give here would ever appear in speech.

Extensive examples will be needed. It should be noted that there are no examples of constructions with complex logically shared final termsets in the NB3 corpus.

A predicate1 is either a predicate2 or a predicate2 followed by an AGE connective followed by a predicate1. Notice that there is no provision for adding termsets shared via AGE connectives, and also that these highest level afterthought connectives group to the right.

A predicate is a predicate1 or an identpred (one of the identity predicates listed above in the lexicography section; note that we intend to allow these last to be prefixed with **nu**).

6.3 Clauses, arguments and term lists

In this term we do the constructions which culminate in terms (arguments and modifiers) and term lists.

6.3.1 Serial names and the false name marker problem

A name word refers to either a consonant final name word or an acronymic name. We already know that such words must be followed by pauses.

The words **la**, **hoi**, **ci**, **hue**, **liu**, **gao** are the "name markers". A name word must be preceded either by a pause or by a name marker.

An occurrence of a string identical to a name marker word in a name is called a "false name marker" if what follows the apparent name marker word is itself a well formed name word.

Complex name constructions are supported (serial names). A serial name begins with a name word, followed by a series of items of the following sorts (each of which will begin with at least a space):

1. `ci` (possibly preceded by a free modifier) followed optionally by a pause followed by a name word, as in **Pierr ci Laplas**

2. `ci` (possibly preceded by a free modifier) followed optionally by a pause followed by a predunit; this may not be followed immediately by an item of the next type (an unmarked name word), as in **Djan ci Blanu**

3. an unmarked non-acronymic name word containing no false name markers, as in **Djan Braon** (never preceded in a serial name by a predunit).

It should be recalled that a name word is always followed by an explicit pause, except when it is followed by end of text, terminal punctuation, a space followed by **ci** or a space followed by another name word (commas are permitted but not required in the latter two contexts). The last special cases are motivated as we can now see by the structure of serial names. In the special cases, there is a pause at the end of the name word even though it is not expressed by a comma.

We add some remarks about the general problem of false name markers. The issue is whether we can tell where a name word starts. The end of a name word is always detectable as an explicit pause (or terminal punctuation, or a space before **ci** or another name word). The problem is ensuring that we can recognize the beginning of a name word. The key to our solution is that the parser will only attempt to read a name word starting in very precisely defined positions: immediately after an explicit pause, or a name marker word, or another name word already read. Moreover, name words only appear in quite specific grammatical contexts (this was enforced by eliminating unmarked vocatives (addressing John as just **Djan** rather than **hoi Djan**) which made it possible for name words to be free modifiers capable of appearing almost anywhere), and by making some further technical modifications in how name words can appear in other grammatical constructions.

Where any name marker appears followed (with an optional intervening comma pause) by something which can be read as a serial name, this is the actual parse which will be produced. If this parse is not intended, perhaps the speaker should pause somewhere (an unintendedly false name marker can be made a true one by putting a comma pause after it).

6.3.2 Arguments (including subordinate clauses)

These are the pronouns and noun phrases of Loglan.

We begin with some preliminaries.

A gap, we remind ourselves, is either a significant comma pause, or **gu** with optional comma pauses before and/or after it.

A laname is **la** followed by an optional comma pause then a mandatory serial name, then an optional gap.

A vocative is either **hoi** followed by an optional comma pause followed by a serial name followed by an optional gap, or **hoi** followed by an optional free modifier followed by a descpred followed by an optional gap, or **hoi** followed by an optional free modifier followed by an argument followed by an optional gap, or just **hoi** followed by an optional gap. Notice that all vocatives are marked with **hoi**. A serial name by itself is not even an utterance. Notice that vocatives

are themselves free modifiers. [NOTE: I am considering adding **sia**, **siu**, **sie** as additional vocative markers. Could register cmapua be vocative markers?]

We now present a series of classes which are sorts of argument.

A descriptn is one of the following sequence of kinds of descriptive phrase (considered in this order):

1. A LE word followed by an optional freemod followed by a descpred. This is very basic: **le mrenu**, **le cmalo hasfa**, etc. This is guarded against being an initial segment of a LANAME (so that something like **la Hasfaran** really is read as a LANAME, not as **la Hasfa**, leaving the ran dangling).

2. A LE word followed by a mex (mathematical expression) followed by a descpred. Freemods are allowed in both medial positions. **Le to mrenu** is a simple example.

3. A LE word followed by an optional freemod followed by an arg1 (a particular flavor of argument to be revealed) followed by a descpred. **Le le hasfa, mrenu** is an example. This is a sort of possessive.

4. A LE word followed by a mex followed by an arg1a (another subtle flavor of argument; a difference to be noted is that a pronoun is an arg1a but not an arg1) [with freemods insertable in medial positions]. An example is **Le to le mrenu** We can have **Le to mi**, whereas we cannot have **Le mi hasfa** under the previous heading (though we do have **Lemi hasfa**, with the right meaning, because **lemi** is a LE word).

5. **ge** followed by a mex followed by a descpred. I need to firm up my understanding of what this case is for.

An arg1 is one of the following quite long laundry list of noun phrase constructions (tested for in the order given by the parser):

1. A LEFORPO word (this class includes LE and the NI cores) followed by a PO word followed by an uttAx, optionally closed with **guo** or a gap. Freemods may appear in medial positions.

2. A LEFORPO word followed by a PO word followed by a sentence, optionally closed with **guo** or a gap. Freemods may appear in medial positions. Note in both of the first two cases that these do not have a PO-initial predunit as a component, though it looks like it is there. We avoid parsing these constructions to include such a predicate to avoid having to close these constructions twice (once for the predicate and once for the argument).

3. **lio** followed by either a descpred or a term or a mex (tested for in that order) closed optionally by a gap.

4. a foreign name starting with **lao** (details under lexicography).

5. a laname (described in preliminaries above). It is important to note that laname is preferred to descriptn. As above, **la Has'faran** is a name.

6. a descriptn, optionally followed by a non-pause freemod, followed optionally by a serial name, with the serial name being marked initially either by **ci** (optionally flanked on either side by explicit comma pauses, or just by an explicit comma pause, in which case the initial name word in the serial name should be non-acronymic and contain no false name markers, all closed optionally by a gap.

   The construction without a name is just the very common **le cmalo hasfa**. With a name, we have such things as **le blanu, Djan** or **le blanu ci Djan**, "Blue John". A practical example of this is in forms of address: **Le surpoi, Djonz**, "Lord Jones". I think Mr, Mrs., Miss should be implemented in this way (I am not saying that this was intended, but it is a clear use of this construction).

   The requirement of a explicit comma pause before the optional name here (when **ci** is not used) is I believe new, a feature of the general solution of the false name marker problem.

7. A word quoted with LIU or NIU, or a letter quoted with LII. See above.

8. A LIE strong quotation. See above.

9. A LI quotation. See above.

The slightly richer class arg1a of arguments consists of the following kinds (tested for in this order):

1. a DA pronoun

2. a TAI letteral pronoun

3. an arg1

4. **ge** followed by an optional freemod followed by an arg1a. I need to understand the use of prefixing **ge** here.

An arg1a of any of these shapes may further include a following non-pause freemod.

Note that this class adds in the pronouns. This was already important in understanding descriptn above.

We now introduce argument modifiers (subordinate clauses).

An argmod1 (atomc subordinate clause) consists of an optional **no** of negation (not currently allowed to be followed by a freemod – should I allow this?), then one of the following:

1. a JI word followed by a predicate

2. a JIO word followed by a sentence or uttAx

3. a JI word followed by a modifier (a relative clause)

4. a JI word followed by an argument

Any of these forms are optionally closed by GUI (the word **gui**, optionally attended by commas before and/or after, or a gap).

A construction of the class argmod is a series of argmod1's linked by A logical connectives (NOTE: the gap in the argmod class definition seems to be redundant).

An argument of class arg2 is an arg1a or an arg1a followed by one or more argmods. NOTE: Again, the optional gap in the arg2 rule may be redundant. NOTE: Should there be a medial freemod in the arg2 rule? I think there may be a reason not to have one. NOTE: look carefully at closure of argmods, arg2's.

An arg3 is either an arg2 or an arg2 preceded by a mex (a quantifier) with a medial freemod allowed. e.g., **ra le mrenu**.

An indef1 is a mex followed by an optional freemod followed by a descpred (e.g., **to mrenu**).

An indefinite is an indef1 followed by an argmod. NOTE: as above, should a freemod be allowed medially? There may be a reason not to allow this.

An arg4 is a string of (possibly mixed) arg3's and indefinites linked by **ze**: this forms mixed arguments. This is a distinct grammatical usage of **ze** from the one as an instance of CA.

An arg5 is an arg4 or an arg4 forethought connected to an argx (this class is described soon below): this form is a KA word followed by an arg4 followed by a KI word followed by an argx, with medial freemods allowed.

An arg6 is an arg5 possibly modified by a positional or semantic case tag, **lae** or **lue**, or the operator **ie** (interrogative which)). Repeated modifications are supported. Medial freemods are allowed.

An argx is a possibly multiply negated arg6 (negation being achieved as usual by prefixing **no** followed by an optional freemod.

An arg7 is a chain of argx's linked by ACI logical connectives. These group to the left as always.

An arg8 is a chain of arg7's linked by A connectives: an arg8 is further constrained not to begin with the cmapua **ge**.

An argument is a set or list form (to be discussed later [very possibly not at all in this draft]; this construction is one I have done little analysis of) or a chain of arg8's linked by AGE connectives, optionally followed by a GUU followed by an argmod (allowing attachment of a subordinate clause at the very top level). NOTE: I am very suspicious of the fact that explicit set and list forms are not allowed to enter into any complex argument constructions. I expect that when I do pay attention to them I will entirely change their role in the grammar, much as I did with (PO sentence) predicates. I suspect that AGE connectives should group to the right as AGE predicate connectives certainly do.

NOTE: the ability to attach subordinate clauses only to low complexity arguments or at the very top level may be a limitation.

NOTE: do we want to be able to forethought connect subordinate clauses?

NOTE: I have a general concern about where closures of argument constructions are or are not needed. I note that basic description constructions do not have closures at all (different from the situation in **la Sorme Lengu**), but some complex constructions do close. [I do find that descriptn can be closed with a gap because descriptn gap is a case of arg1]. I have an overall impression that closures of constructions involving argmods would benefit from an overhaul.

6.3.3 Explicit set and ordered list forms (omitted for now)

Omitted for the moment. These require analysis; they are not first-class citizens of the language in the grammar as presented in trial.85.

6.3.4 Modifiers = relative clauses, prepositional phrases

A tense/location/relative clause (class mod1) consists of a PA word followed by an argument, optionally closed with GUU (either **guu**, optionally flanked on one or both sides with a comma pause, or a gap), or a PA word optionally followed by **guu** (a gap does not work in the absence of an argument). The second option gives a relative clause which can be distinguished from a tense (a PA word included in a markpred).

A kekmod is a forethought connected modifier: this consists of zero or more **no**'s, followed by a KA word, followed by a top level modifier, followed by a KI word, followed by a modifier of class mod (defined immediately below). Medial freemods are allowed.

A mod is either a mod1, or a mod1 prefixed with one or more negations, or a kekmod.

A modifier (top level) is a mod or a chain of mods linked (left grouped as usual) by A connectives.

6.3.5 Terms, term lists, and termsets (including linkargs)

A term is an argument or modifier.

There are two kinds of argument lists, the loose lists which are conglomerations of terms with no explicit operator optionally closed with GUU and the tightly bound lists built with JE/JUE.

A construction of class terms is a sequence of one or more terms with optional medial freemods. The lists of terms in the definitions of the basic sentence classes are of this grammatical class (termsets appear internally to predicates only).

A termset1 is a construction of class terms optionally closed with GUU (either **guu** optionally flanked with explicit commas on one or both sides or a gap) or a forethought construction: a KA word followed by a termset2 followed by a KI word followed by a termset1 (medial freemods allowed). NOTE: should I include optional negations here?

A termset2 is a sequence of termset1's linked by A connectives, with medial freemods allowed.

A termset is one of the following:

1. an item of class terms followed by **go** followed by a barepred. The barepred modifies the predicate to which the termset is attached; this is a weird but I think useful maneuver.

2. a termset2

3. GUU = the word **guu** (optionally flanked by pauses on one or both sides) or a gap. This null case is actually quite useful where logically shared termsets are in play.

Termset is an important class, having been introduced earlier as a black box internal feature of the predicate classes.

We now consider the tightly bound lists (culminating in class linkargs).

A juelink is **jue** followed by a term.

A links1 is a sequence of juelinks optionally closed with GUE (either the word **gue**, optionally flanked with explicit pauses on one or both sides, or a gap).

A links is one of the following:

1. a links1

2. a KA word followed by a links followed by a KI word followed by a links1 (medial freemods allowed).

3. a sequence of items of one of the previous two types linked by A connectives (left grouped as usual).

A jelink is **je** followed by a term.

A linkargs1 is a jelink followed by a links optionally closed with a GUE (NOTE: the gue in this rule might be problematic, forcing double closures in some situations).

A linkargs is one of the following:

1. a linkargs1

2. a KA word followed by a linkargs followed by a KI word followed by a linkargs1 (medial freemods allowed).

3. a sequence of items of one of the previous two types linked by A connectives (left grouped as usual).

The idea is that in these tightly bound argument lists first arguments are attached with **je** and second and subsequent arguments are attached with **jue**, which often reduces the occasion for explicit closures. Recall that the class linkargs enters into the construction of predunits.