Loglan Grammar Proposal, 2018

M. Randall Holmes

4/14/2018

1 Introduction

In this document, I will attempt to describe the grammar proper (phonetics and lexicography being handled in previous documents) which is implemented in my provisional Loglan parser(s). I hope that the Academy will approve this (along with the previous two documents) as the official Loglan grammar; in any case, this describes the language parsed by my parsers, and a language which is in practice very close to the 1989 Loglan language as modified by decisions taken by the previous Academy in the 1990s. My intentions in my revisions have been to make the language work, not to make far-reaching changes. My intentions are conservative.

2 Sentences, Free Modifiers, and Utterances

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

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

2.1 The basic SVO statement

The most basic Loglan sentence form consists of a subject followed by a predicate: we call this a basic SVO statement. Of course, a lot of complexity

is hidden in these words, and there are some additional optional components.

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

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

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

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

Here is a simple example:

La Djan, cluva la Meris

And another:

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

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

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

We give an example

La Djan, gio le bakso ga donsu le Meris

The same effect is achieved by

La Djan, zue le bakso ga donsu la Meris

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

A "sentence" like

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

is permitted in 1989 Loglan but forbidden by our grammar.

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

2.2 Subject-final statements and observatives

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

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

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

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

The second class of gasents consists of a tense word, followed by an untensed "verb phrase" (no objects), followed by **ga** followed by a term list (which may optionally be of the form (subject + **gio** + term list), but use of **gio** is not required).

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

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

Na crina

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

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

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

2.3 Sentences

The Loglan unit sentence is one of the following:

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

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

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

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

forethought connected sentence: A forethought connected sentence consists of a KA core followed by a Loglan sentence, followed by **ki** or **kinoi** followed by a unit sentence or a sentence with head terms. This kind of sentence may optionally be multiply negated with initial occurrences of **no**. This is a species of unit sentence.

This is revised from a rule carried forward from 1989 Loglan until very recent versions of my prover which allowed the final component of a forethought connected sentence to belong to the very general class of utterances `uttA1`, which contains many sorts of sentence fragments normally uttered as answers to questions, as well as the two kinds of sentences indicated. To our minds, this situation was bizarre. In

parsing the Visit to Loglandia we found a number of grammatical errors not detected because of the use of `uttA1`.

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

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

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

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

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

in which the object is first, with the caution that one again has to attend to the fact that any fronted term list applies to all sentences in the logically connected sentence to which it is fronted.

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

2.3.1 Quantifier Scope

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

Where two indefinites have the same scope, the outer quantifier is determined by which one appears first in the utterance. **Raba cluva be** means that for every x there is a y such that x loves y. **Be nu cluva raba** means that there is a y such that for every x, x loves y (notice how use of the converse enables us to draw this distinction, much as it can in English). There is a qualification: where the apparent first occurrence of an indefinite has an appended subordinate clause (with JI or JIO) which contains an occurrence of a pronoun referring back to the indefinite, the first such pronoun counts as the first occurrence of the indefinite (this effect is recursive in nested subordinate clauses). **To mrenu jio te fumna ga cluva mei** refers to three men who are loved by *the same* three women, while **To mrenu jio mei nu cluva te fumna** are two men each of whom are loved by three women (possibly different women for each man).

A prenex quantifier string ending in **goi** has as its scope exactly the sentence with head terms for which it is the head term list. Modifiers in the prenex list are just modifiers. The items in the quantifier list must be indefinites. They may be qualified with subordinate clauses with **ji** or **jio** to indicate restricted scope (use of **jio** allows more complex restrictions than

the usual logical notation). Components of the form **ra bua** (where **bua** is a predicate bound variable) are read as quantifying over predicates rather than being bounded to an indefinite predicate. **Ra mrenu goi mei cluva mei** quantifies over men (all men love themselves). **Ra bua ra ba ra be goi ba bua be, ico be bua ba** makes the unlikely assertion that all relations are symmetric: it quantifies over all predicates rather than all things which are **bua**. This is a dodge to accommodate second order logic.

2.4 Note on right closers

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

2.5 Free Modifiers

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

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

The other flavors of free modifier are:

1. Words of class UI or NOUI (attitudinals and negative attitudinals).
2. Spoken smilies: **soi** followed by a descriptive predicate (class `descpred`) optionally closed with the right closer **guea** or **gu**.

3. One of the register words of class DIE, optionally negated by being prefixed with **no**.

4. A parenthetical utterance `kie` + utterance + `kiu`. The utterance must be a well formed Loglan utterance. It may be set off (after `kie`, before `kiu`) either with commas or with a pair of parentheses.

5. inverse vocatives, to be discussed just below.

6. vocatives, to be discussed just below.

7. items of class JO, the word **jo** optionally preceded by a digit, which has the effect of putting a number of words preceding the JO item in "scare quotes". The number of words "quoted" is indicated by the digit if it is present, and otherwise is one.

8. ellipses `...` and double hyphens `--`.

A freemod can always optionally include another freemod.

2.5.1 Vocatives

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

1. a name (possibly set off from the vocative marker by a comma-marked pause),

2. or a descriptive predicate (`descpred`) which may optionally be closed with the right closer GUEA (`guea` or `gu`) optionally followed by a name, which may be marked with **ci** and must be so marked if it contains a false name marker.

3. or an argument without a case tag (**argument1**), possibly set off from the vocative marker by a comma-marked pause and possibly closed with the right closer GUUA (**guua** or **gu**),

4. or a foreign name (alien text): in this case the vocative marker must be **hoi** and the alien text must be enclosed in double quotes. The double quote requirement is to prevent accidental acceptance of buggy Loglan text as a foreign name.

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

2.5.2 Inverse Vocatives

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

1. A name.

2. A descriptive predicate, possibly closed with the right closer GUEA, possibly followed by a name which may be marked with **ci**. This is just as in the vocative construction, except that a free modifier is allowed after **hue**.

3. A statement possibly closed with the right closer GIUO (**giuo** or **gu**). This allows free modifiers of the form "said John", as it were.

4. An argument without a case tag (`argument1`), possibly closed with the right closer **guu** or **gu**.

5. A foreign name (alien text), which must be double quoted.

2.6 Utterances

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

1. A unit sentence.

2. A sentence with head terms (notice that a mere logically connected sentence is not of this class).

3. tightly linked term lists with **je** and **jue** (classes **links** and **linkargs**, described below).

4. subordinate clauses (class **argmod**).

5. a term list. NOTE: why are termsets not in this class?

6. a brief answer of class `uttA` (either a logical connective or a quantifier/number).

7. An occurrence of **no** as a word.

A general answer may optionally be suffixed with terminal punctuation.

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

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

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

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

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

Notice that we have arranged for a logically connected sentence such as

da redro, ice da blanu

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

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

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

1. an I connective, optionally followed by terminal punctuation and optionally further followed by an I-connective initial utterance.

2. a free modifier other than a pause (optionally prefixed with an I connective and optionally suffixed with terminal punctuation, and further optionally followed by an utterance of any form). This is essentially the only case where a free modifier might appear initially.

3. an `uttF` followed by an IGE connective followed by an utterance.

4. an `uttF` optionally prefixed with an I- or ICA connective, followed optionally by an I connective-initial utterance.

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

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

3 Predicates

Our treatment of predicates may conveniently be divided into treatment of verb phrases (which do not have termsets attached, though they may have tightly bound term lists built with **je** and **jue** attached) and predicates per se, which may have termsets attached.

3.1 Verb Phrases

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

3.1.1 Simple Verb Phrases

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

1. A predicate word, possibly preceded by a conversion or reflexive operator of class NU.

2. A foreign predicate or onomatopoeic predicate (**sue** or **sao** followed by alien text).

3. A possibly complex descriptive predicate turned into a `predunit` using the initial marker **ge** and (optionally) the final marker **geu** or **cue** (**cue** is an older form of class GEU: this is not a right closer class and does not have **gu** as a possible shape). This takes two forms: the basic form is **ge** + **descpred** + (optionally) GEU and the converse or reflexive form NU + **ge** + `despredE` + (optionally) GEU. I need to look into the reasons why slightly different classes of descriptive predicates are allowed here. Grouping with **ge**...(GEU) is useful in expressing complex metaphors precisely.

4. A predicate built from an argument without a case tag: `me` + `argument1` possibly closed with the right closer MEU (**meu** or **gu**).

5. An abstraction predicate: this is one of the abstractors `po`, `pu`, `zo` followed by a sentence (unit sentence, logically connected sentence or sentence with head terms) possibly closed with the right closer GUO.

Alternatively, forms of the abstractors suffixed with **(z)a, e, (z)i, o, (z)u** may be used, in which case forms of the right closer with the same suffix may be used: this allows closure of several nested abstraction predicates (or abstract descriptions) with a single right closer.

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

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

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

A `predunit3` is a `predunit2` possibly followed by a tightly bound term list built with **je/jue**.

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

A further "unit" predicate is the forethought connected predicate, which consists of an optional prefix of one or more **no**'s, followed by a forethought connective (KA) followed by a predicate (of the most general form), followed by a class KI word completing the forethought connection, followed by another predicate, optionally closed with the right closer **guu** or **gu**. Note that the **guu** closure, introduced late, removed the reasons for the rule forbidding forethought connected predicates as heads of predicate modifications ("metaphors"). This is an arbitrarily complex predicate construction which is treated as a simple verb phrase because it is suitably packaged.

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

3.1.2 Complex verb phrases formed with predicate modification and tight logical connection

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

I am referring to the process which JCB calls "metaphor" as "verb modification".

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

A `despredC` is a series of `despredB`'s with no intervening structure word, where a `despredB` is either a `despredA` (the usual situation) or a construction of the form (`cui` + `despredC` + CA + `despredB`): this is a verb modification construction, binding more loosely than the construction with `ci`, with the further ability to logically connect `despredC`'s on the left with its terms (repeatedly), guarding the logical connections with the left marker `cui`. This particular construction is actually used only internally to `despredB`.

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

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

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

A sentence predicate (or sentence verb phrase) is either a `despredE` or a `despredE` followed by `go` followed by a barepred (a class of sentence predicates

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

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

3.2 Predicates

4 Arguments, Modifiers, and Term Lists