

A proposed complete description of TLI Loglan phonetics

Randall Holmes

3/10/2018: added motivational section at end. Dec 4, 2017 9 pm, remark on phonemic status of syllable breaks and an additional form of djifoa with illegal syllable break. Added a comment about default placement of syllable breaks in strings of consonants. A few additional tweaks 1 pm on the 4th. Added the traditional notation for consonant and vowel patterns. Added a couple of footnotes; main text is frozen except for typo fixes. Name word section redone, grumble. Hopefully frozen again. Dec 11 added footnotes about the new PEG implementation, and also a minor proposal to support names for **y**. Dec 12 added appendix on alien text and quotations. Dec 16 updated with some refinements about resolving vowel sequences. 12/17/2017 added description of tests for the start of a predicate.

I wrote this document recently as part of a larger document with a different purpose. It struck me that it could be used as a framework for presenting all of my current phonetic proposals to the Academy in the context of a complete description of the proposed phonetics of Loglan. An interesting point is that there are not actually very many new proposals! This goes along with my basically conservative approach to what we have inherited. I'm hoping that perhaps this can be Approved. Comment by all Loglanists is encouraged!

I think that all of this works. Some of it is quite baroque and could be improved in various ways (some of it perhaps *should* be improved), but my present mission is to get an official view of what the current state of affairs is. Essentially everything here is about phonetics and word forms: there are

allusions to other perhaps controversial features of my provisional grammar (such as converting acronyms to names) but these have no bearing on the phonetics.

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

1 Sounds

1.1 Vowels

The Loglan vowels are the *regular vowels* **a,e,i,o,u** and the *irregular vowel* **y**. The pronunciations of the regular vowels are typical Continental European (not English!) pronunciations. The sound of **y** is officially the schwa, but we think there is something to be said for it being another sound easily distinguished from the regular vowels, such as the *oo* in English *look*. We have also considered *ö* and the Cyrillic letter that looks like *bl*² as implementations of Loglan **y**.

1.2 Consonants

The Loglan consonants are **b,d,f,g,h,j,k,l,m,n,p,r,s,t,v,z**. The pronunciations of these are standard European pronunciations, except that **c** is English *sh* and **j** is the corresponding voiced sound found in English *azure*. **g** is always “hard”. **h** has an alternative pronunciation as *ch* in Scottish English *loch* when final in a syllable³ [this is NEW: **h** does not occur in syllable final position in 1989 Loglan]. **n** is pronounced as *ng* in English *sing* when it appears before **k,g** or syllable final **h**. NEW: **q,w,x** with odd pronunciations are eliminated: this was already mostly carried out in the 1990’s.

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

²It is *not* easy to insert this in LaTeX!

³I accept an amendment from John Cowan allowing the alternative pronunciation of **h** in all contexts.

1.3 Vowels pretending to be consonants

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

1.4 Consonants pretending to be vowels

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

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

2 Diphthongs and vowel grouping

There are some mandatory and some optional pairs of vowels which form diphthongs⁴, which can serve as the vowel component of a single syllable.

2.1 Mandatory diphthongs

The mandatory diphthongs are **ai**, **ao**, **ei**, **oi**. The pronunciations are as one would expect from the values of the vowels, except that **ao** is as in English *cow*. These vowel pairs are not regarded as mandatory diphthongs when followed by another **i**.

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

2.2 Optional diphthongs

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

⁴I looked this word up in the dictionary, and it does seem to refer to vowel combinations in a single syllable, which is what is desired.

vowels is not pronounced as a diphthong, one may flow into the other without pause, or a glottal stop (not expressed in writing except possibly indirectly by a hyphen or stress mark) may be inserted (NEW: we do not allow the glottal stop to be an allophone of the pause, as earlier versions of Loglan did; in Lojban they require insertion of what we would write **h** in non-diphthong vowel pairs).

2.3 Grouping long streams of vowels in predicates and names

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

1. group the first two vowels and continue if they make up a mandatory diphthong (or **ii**, **uu**⁵).
2. pronounce the first vowel as a single syllable and continue if the second two make up a mandatory diphthong (or **ii**, **uu**) and the first two do not (or in the cases **aii**, **eii**, **oii**).
3. optionally, take the first vowel as a syllable and continue or group the first two vowels and continue if they make up an optional diphthong and the previous condition does not hold; the parser will always take the second alternative.
4. pronounce the first vowel as a single syllable and continue, if none of the previous conditions hold.
5. by “continue”, we mean “apply the same set of rules to the remainder of the stream of vowels”.

This is NEW, not the same as the rule in 1989 Loglan (given there only for names) but it appears to have similar effects in practice.⁶

⁵Special treatment of doubled **i** and **u** is designed to avoid formation of **i-i** or **u-u**, which would trigger the double vowel stress rule.

⁶The rule in 1989 Loglan calls for right-grouping, even of very long streams of vowels, which is psychologically incredible, as one cannot necessarily tell at the start of a long

2.4 Grouping long streams of vowels in structure words

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

2.5 Vowel pairs with optional grouping revisited

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

2.6 Doubled vowels and stress

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

3 The Loglan syllable

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

We now discuss the Loglan syllable. Each Loglan syllable is either unstressed, stressed, or emphatically stressed. Syllables may be separated by a hyphen. A marker ' of stress or * of emphatic stress may terminate a syllable⁷: a stress marker may not be followed by a hyphen, as the stress marker itself serves the

stream of vowels whether it is of odd or even length. I think the reasons why JCB wanted to right-group are captured by the second clause of my algorithm.

⁷It is important to note that it is the end of the syllable, not the vocalic unit of the syllable, which is marked with the stress.

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

Note that a syllable without a stress marker is not necessarily unstressed: we simply have not committed ourselves.⁸

3.2 Every syllable has a vocalic unit

Each Loglan syllable contains a vocalic unit, which is either a single vowel⁹, a diphthong pronounced as such, or a doubled continuant.

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

Each Loglan syllable consists of up to three parts.

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

⁸Many syllables in actual Loglan text can be recognized as definitely unstressed from context, e.g., most syllables in predicate words, but there is no marker for this so far.

⁹the irregular vowel **y** can be a vocalic unit, though its distribution is limited; it can occur freely in names and as a “phonetic hyphen” in complex predicates, but not at all in *cmapua* or borrowed predicates: some changes allowing its use in *cmapua* in limited contexts are implemented in the new PEG.

pair.¹⁰

2. The second part is the mandatory vocalic unit.
3. The third part, which is optional, consists of one or two final consonants. There is a list of impermissible medial pairs and a list of impermissible medial triples.¹¹ A final consonant may not be followed by a vowel and may not stand at the beginning of an impermissible medial pair or triple (regardless of whether the other consonants are in the same syllable, and ignoring stress marks and hyphens). A continuant pair is an impermissible medial pair: a final consonant may not participate in such a pair, even with a syllable break intervening.

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

In placing syllable breaks in the absence of an explicit hyphen or stress mark, the rule is followed that a final consonant shall not stand at the beginning of a legal syllable, except in the case of a first final consonant at the end of a syllable whose initial consonant group consists of a single consonant and whose vocalic group is a single regular vowel (the general idea is that the syllable break is usually placed as early as possible, except that a CVC syllable is read in preference to a CV syllable where possible).¹²

¹⁰The 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**

¹¹The impermissible medial pairs consist of all doubled consonants, any pair beginning with **h**, any pair both of which are taken from **cjsz, fv, kg, pb, td**, any of **fkpt** followed by either of **jz, bj**, and **sb**.

There is a list of impermissible medial triples as well, consisting of **cdz, cvl, ndj, ndz, dcm, dct, dts, pdz, gts, gzb, svl, jdj, jtc, jts, jvr, 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.

¹²The old parser behaved differently in this respect in names, allowing no final consonant to stand at the beginning of a legal syllable in the absence of an explicit syllable break. This wasn't a difficulty; there are reasons for the more nuanced rule to be applied in predicates. In complex predicates, any explicit syllable break placement should agree with djifoa structure. In borrowed predicates, the placement of explicit syllable breaks was governed by some rather mysterious rules which had the effect of preventing one from obtaining a

3.4 Words must resolve into syllables

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

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

3.5 Notation for consonant and vowel patterns

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

4 Word Forms

4.1 Word forms enumerated

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

4.2 Pauses and word boundaries

Words (in the phonetic sense) end at whitespace, at a comma or mark of terminal punctuation (periods and some other punctuation marks; we do not give the list here, and in general remark that there are some subtleties of punctuation rather than phonetics in the provisional parser which we do

legal borrowing predicate by moving syllable breaks in illegal complex predicates in a way which violates djifoa structure: below we give a different description of which borrowing predicates should be excluded for this reason, which we have implemented in the parser, and which allows freer placement of explicit syllable breaks.

not cover here), or sometimes without any explicit indication at all (where phonetics are sufficient to recognize where one word ends and another begins).

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

Whitespace is sometimes an explicit pause and sometimes a word boundary which is not marked by any actual phonetic feature. Where whitespace does not appear, one should not pause. Where a pause is allowed at whitespace, a comma should always be permitted (The old parser LIP does not always support this, but we regard this as debugging, not a novelty). There are situations where whitespace is allowed due to a word break but an actual comma pause would change the parse (and so in speech such a whitespace is not expressible as a pause).¹³

Vowel initial words are always preceded by a pause if they are not at the start of a text or utterance. Consonant final words are always followed by a pause if they are not at the end of a text or utterance. Thus, whitespace preceded by a consonant or whitespace followed by a vowel must represent an actual pause. So we also regard whitespace preceded by consonants or followed by vowels as an explicit pause. It appears to be NEW that we must pause before the first in a stream of VV words, but it is also clearly necessary, as experiments with phonetic transcripts have revealed.¹⁴

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

It is possible to resolve a stream of Loglan phonemes into words unam-

¹³This last sentence is true in the present state of my parser (because of one case in the way it handles name markers) but this feature is not in this proposal, and I am considering disabling it. No (added later) there are other instances of whitespace which are permitted to be written but cannot represent a pause, in connection with the handling of the legacy forms of the APA connectives, another pain in the language, and quite another story.

¹⁴The new PEG implementation supports the traditional requirements that pauses at the end of a serial name and pauses before a logical connective must be actual comma pauses. Some logical connectives are consonant-initial: there is a purely phonetic description of the front of a logical connective in the PEG.

biguously, with the qualification that the resolution of streams of Loglan grammatical particles is actually done by the grammar proper. We indicate how to do this.

4.3 Structure words (cmapua)

A structure word (in the phonetic sense: some phonetic structure words are actually semantically names or predicates) is a word which resolves into a stream of V, VV, CV, CVV, and Cvv-V units (where vv denotes a diphthong). These words are called **cmapua** in Loglan. The CVV units do not have to be syllables (there is no requirement that the VV be a diphthong, or that it be pronounced as such if it is an optional diphthong). A V unit can only appear initially; if any unit in a cmapua is of the shape VV, all units are of the shape VV. We recall the rule that one must pause before a word which begins with a vowel, so one must pause before an initial V- unit; it is not necessary to pause between VV units in a phonetic word made up of such units, but it is necessary to pause before a lone VV unit or the first in a stream of such units (the necessity of pausing before the first in a stream of VV units seems not to have been recognized in NB3). Except in the case of the end of a string of VV units, it is impossible to recognize the ends of individual words in a stream of cmapua phonetic units on phonetic grounds alone: the grammar proper allows us to resolve streams of unit cmapua into words, but for phonetic purposes we may regard streams of VV units and streams of non-VV units as “words”.¹⁵

To support names for the letter **y** (and perhaps add a little cmapua space), we propose to allow the irregular monosyllabic diphthongs **iy** and **uy** in CVV (but not Cvv-V) cmapua units. We further propose to allow **y** to be a V djifoa unit (in compounds). This makes the name **yfi** of **y** legal.

It is part of the definition of a cmapua unit that it cannot be an initial segment of a predicate (either complex or borrowed): a more precise statement of this is that a cmapua unit may not be followed without an intervening explicit pause by **y** or by **CyC**, and may only be followed without an intervening explicit pause by **CC** if the **CC** itself is initial in a predicate word (this last condition requiring lookahead!). This indicates the (sometimes rather subtle) way that the end of a phonetic structure word at the

¹⁵There is a further subtlety that certain streams of unit cmapua which are broken by explicit pauses may be recognized semantically as words of class PA or NI: this is a semantic rather than a phonetic proposal, not discussed here

beginning of a predicate word is recognized; how to recognize the end of a structure word at the beginning of a name word is covered in the next section (and may require considerable lookahead).

There is a precise set of conditions under which an apparent *cmapua* unit cannot be part of a *cmapua* word because it is the start of a predicate word (if it is anything legal). It is further the case that the start of a borrowing or complex not starting with an initial pair or triple of consonants must pass one of these tests (an initial CV^n does not have to be a legal *cmapua* unit in a predicate, but does have to satisfy one of these conditions):

1. If the unit is followed without pause by two or more consonants, where the consonant group is not an initial group of consonants or is a consonant or initial group of consonants followed by a syllabic pair.
2. If the unit stands at the beginning of a copy of a $CVV\mathbf{y}$, $CVC\mathbf{y}$, or $CVCC\mathbf{y}$ *djifoa* (see below for discussion of *djifoa*).
3. If the unit is finally stressed and immediately followed without pause by more than one consonant and is not a V unit, $Cv\mathbf{v}\text{-}V$ unit, or the head of a stream of VV units followed by ccV (this last cluster of cases is simply illegal: no predicate can start in that way for technical reasons to do with borrowing *djifoa* falling apart).
4. If the unit is followed without pause by an initial group of consonants which is then followed by V or VV followed by whitespace or non-syllable-break punctuation or by a stressed V followed by a single V not in a diphthong.

There is an exception: a V or $Cv\mathbf{v}\text{-}V$ *cmapua* unit or a stream of VV units may be followed by ccV (the cc being an initial pair) without intervening pause if not finally stressed. We note that the parser applies these exact tests rather than looking ahead to see if a predicate word actually starts at the beginning of the apparent unit under consideration.

Where the final unit of a phonetic *cmapua* immediately precedes a predicate word and is stressed, it must be followed by a pause: if the stress is marked in writing and the following predicate is consonant initial, the pause must be explicitly written as a comma. We are pleased with the fact that this rule, going back to the beginnings of Loglan, can be expressed in our orthography and is enforced by our parser; it is quite invisible to LIP.¹⁶

¹⁶We note a subtle point about the articulation of acronyms in Loglan: these are seman-

4.4 Name words and the false name marker problem

A name word is a stream of Loglan syllables ending in a consonant followed by whitespace, a comma marked pause, or terminal punctuation. Name words are the only Loglan words which end in consonants. The parser currently requires an explicit comma pause in place of whitespace after a name word in most contexts, though the fact that a consonant followed by whitespace is recognized as an explicit pause suggests that this can be relaxed. The boundary of a name word on the right is readily recognized (the consonant followed by a explicit pause or terminal punctuation). Left boundaries of name words always fall just after either a *cmapua* belonging to a class of “name markers”¹⁷, or an explicit pause. A candidate left boundary for a name word falls just after an explicit pause or just after a phonetic copy of a name marker such that the text between its end and the recognizable right boundary of the name word resolves into syllables (and of course contains no pauses or whitespace). A candidate left boundary is said to be marked if it is just after a name marker or just after an explicit pause immediately preceded by a name marker. The actual left boundary of a name word is the leftmost marked candidate left boundary, if there is one, and otherwise the rightmost candidate left boundary.

We refer to name markers which are candidate left boundaries of name words but are not the actual left boundaries as “false name markers”. Earlier versions of Loglan forbade these, with the odd effect that (for example) **la** could not occur in a name. Lojban still forbids this (and has very few name marker words). Our current grammar of Loglan strongly restricts the situations in which any name word can occur with an unmarked left boundary (for example, the current Academy has banned unmarked vocatives), and

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

¹⁷The name markers are **la**, **hoi**, **hue**, **ci**, **liu**, which must be name markers for one reason or another, **gao** if a proposal to allow this *cmapua* to form letter names from name words is accepted, and in a NEW proposal, the words of social lubrication **loi**, **loa**, **sia**, **sie**, **siu**. If adding the words of social lubrication is thought excessive, we can continue to say **Loi hoi Djan** instead of **Loi Djan**: in fact, the new PEG does not have these words as name markers.

in any such context there is provision for the boundary to be marked with a suitable name marker if the name word happens to contain a false name marker. This is a generalization of rules for handling serial names adopted in the 1990's.

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

The parser will raise an error if it finds a name marker word followed possibly by an explicit pause, followed by text including whitespace but no explicit pause, followed by a name word (recognizable by its right boundary at a consonant). The problem with this is that because of the intervening whitespace, it cannot be the intention of the writer that the name marker sets the left boundary of the name word, but if the utterance were read without pause, the name marker would indeed set the left boundary of the name word.¹⁸ There are uses of name markers in which they are not followed immediately by name words: it is the obligation of the speaker to explicitly pause at some point after such an occurrence of a name marker and before the next occurrence of an actual name word, and Loglan orthography requires this to be indicated explicitly. Note that whitespace before a vowel or after a consonant does suffice, but where this doesn't happen, an explicit comma pause (of the form V, C) may be required. Actually complying with this rule is best implemented by style directives such as "always pause after a predicate name", rather than by attention to this esoteric rule as such.¹⁹²⁰

¹⁸A reader requested an example. **La Farfu ga cluva la Djan** would trigger this problem, and needs to be corrected to **La Farfu, ga cluva la Djan**. Without the explicit pause, it could be read phonetically as **lafar'fuclu'valadjan**, which parses as **La Farfugacluvaladjan**, a single name.

¹⁹A directive which would always work is, "always pause after the first word after a name marker, whether it is a name word or not (and whether you paused after the name marker or not)". When what follows the name marker is not a name word, one need not pause after exactly one word, but one should pause at some natural point before the next name word. An example: in **hoi le farfu je la Rabrrt**, which is bad, it is legal but odd to amend it to **hoi le, farfu je la Rabrrt**, most natural to amend it to **hoi le farfu, je la Rabrrt**, and a sign of last minute panic to amend it to **hoi le farfu je la, Rabrrt**.

²⁰It is further important to note that my parser does something different currently in the situation where a name marker is immediately followed by a pause, which seemed

The rule that false name markers cannot occur in names has already been abandoned (in the 1990s) by TLI Loglan; it still obtains in Lojban, which has very few name markers. Precise definition of what you do with false name markers was hard to think about before phonetic parsing was available. The requirement that names resolve into syllables is NEW (I seem to recall that Lojban does something like this?), and interacts with the definition of a false name marker as indicated above. The addition of the words of social lubrication to the list of name markers is NEW.²¹ The automatic detection of dangerous situations where a non-explicit pause should be made explicit is NEW, and we can report from extensive experience in parsing Alex Leith's Visit to Loglandia that there are straightforward ways to correct problems it detects (and that it really does detect things which are potential problems). We can also report, based on extensive parsing of phonetic transcripts of Loglan utterances, that the rules stated above appear to work: the situations in which a name is unintentionally started too early can be controlled.

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

incredibly clever in the context of the kind of parsing I am doing but which I shall probably revert to the exact behavior described here. A first approximation to what my parser does is that it chooses the leftmost candidate left boundary in which the name marker is not followed by a pause, and otherwise the rightmost candidate left boundary: but further, this is handled dynamically by aggressively reading name words after name markers which are not followed by pauses, but attempting to read something else after a name marker which is followed by a pause, and then reading a name word if this attempt at parsing fails: the effect of this is that the resolution into words depends strongly on non-phonetic features of the grammar, which is a bad thing: the new PEG fixes this, behaving exactly as described here.

²¹The new PEG implementation does not add the words of social lubrication as name markers.

4.5 Borrowed predicates

A borrowed predicate is a stream of Loglan syllables which contains just one stressed syllable, whose vocalic unit is not a continuant pair, the stressed syllable being followed by at least one and no more than two syllables, the first of these, if there are two, having a continuant vocalic unit, and the final one not having a continuant vocalic unit and not ending in a consonant: in other words, it is penultimately stressed, ignoring one possible intervening unstressed continuant syllable, and ends in a vowel. A borrowed predicate contains at least one pair of distinct adjacent consonants; it is permissible for one of these to be in a continuant pair. A borrowed predicate may not contain two successive syllables with continuant pairs nor may it start with such a syllable (or, as noted, end with such a syllable), nor may a continuant pair immediately follow a vowel in a borrowed predicate. A borrowed predicate may not contain a doubled vowel unless the doubled vowel is pronounced as a diphthong. A borrowed predicate may not contain **y**. The part of the borrowed predicate before the first pair of distinct adjacent consonants must have the property that omitting it (and dropping any explicit syllable break between the pair of consonants) will not leave a legal borrowed predicate. A borrowing cannot begin *VccV* with the *cc* permissible initial, and there can be no *ccVV* or *cccVV* borrowed predicates²². A borrowing cannot take the shape of a *cmapua* word other than a *CV* or *CVV* unit followed by *ccV*: this is a generalization of the *VccV* prohibition, and has the same motivation.²³

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

²²The *ccVV* predicates are forbidden so that *CVCccV* complex predicates do not have to be **y**-hyphenated; the *cccVV* predicates would, if allowed, greatly complicate our parsing algorithm for a weird technical reason.

²³What needs to be averted is the possibility of reading a borrowing *djifoa* ending with *ccVy* as a stream of phonetic *cmapua* units followed by a predicate beginning with the *ccVy* *djifoa* instead of the intended borrowing *djifoa*. Both the original *VccV* rule and my new (C)VVV...*ccV* rule forbid some predicate shapes which are actually not problematic (but also not needed).

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

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

4.6 Complex predicates

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

The *djifoa* take the forms

1. CVV (legal syllable forms Cvv or CVV or CV-V)
2. CVC
3. ccV (the cc must be permissible initial).
4. ccVCV (when this is not final, the final V is replaced with **y**). The cc must be permissible initial. The only legal syllable break is ccV-CV.

5. CVCCV (when this is not final, the final V is replaced with **y**). The CC must not be impermissible medial. CV-ccV (if the CC is permissible initial) and CVC-CV are legal syllable breaks. Either break is permitted if the CC is permissible initial.
6. a borrowing predicate (modified to have final rather than penultimate stress if not in final position) with appended **y** (in an unstressed syllable by itself) if not in final position.
7. When attempting to resolve a predicate into djifoa to establish that it cannot be a borrowed predicate, the forms to consider are CVV, CVV**r**, CVV**n** (only when followed by **r**), CVC, ccV, and the five-letter forms in final position only, plus the illegal CV-C, c-cV, c-cVCV (the last in final position, the hyphen indicating an explicit syllable break or stress marker; c-c denotes an initial pair broken by an explicit syllable break). A mechanical resolution into these forms, without any side conditions, shows that a string cannot be a borrowed predicate. We do not need to list forms with syllable breaks C-V because these are explicitly forbidden by our phonetics already.

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

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

Any complex includes at least two djifoa, unless it consists of a single five-letter djifoa (the latter could also be viewed as a separate species of primitive predicate).²⁴

Djifoa appearing in non-final position may have phonetic hyphens appended, which may take the shapes **n**, **r**, or **y**. Only one phonetic hyphen can be appended to any djifoa. The consonant hyphens can only be appended to CVV *cmapua*. A phonetic hyphen **n** appears only when followed by **r** in the next djifoa. A phonetic hyphen is never final in a complex predicate nor will it follow a five-letter djifoa. A phonetic hyphen **y** is always unstressed, and appears by itself in a syllable or in a **Cy** syllable in the context CV-**Cy** of a CVC djifoa (CVC-**y** also being allowed; the **y** in a five-letter djifoa in

²⁴The PEG grammar also thinks that a borrowing is a one-complex djifoa, so it classes all predicates as complexes!

non-final position may participate in a **Cy** or **ccy** syllable, which is also always unstressed). A borrowing djifoa is always preceded by **y** if not initial (the **y** will be a hyphen or a constituent of a five-letter or borrowing djifoa) and includes an appended **y** if not final.

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

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

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

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

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

Nothing in the description of complex predicates is new, though all language about borrowing djifoa comes from decisions taken in the 1990s after the 1989 edition of *Loglan 1*. Considerations about explicit stress markers and syllable breaks are new but consistent with the logic of complex predi-

cates as already defined: we do not want an illegal complex to become a legal borrowing by moving a syllable break so that it doesn't conform with a djifoa boundary. We expect that directly implementing the resolution procedure described above will allow us to relax some rather odd rules about placement of explicit syllable breaks which are currently enforced in borrowings under our parser.

5 Appendix: Alien Text and Quotations

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

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

The format for alien text is essentially that allowed to follow **lao** in 1990's Loglan (originally the constructor for Linnaean names: Steve Rice observed correctly that it was better thought of and used as a general foreign name constructor), though the **y**, to be inserted at whitespace was left unexpressed in writing in the original proposal. We require it to be expressed in the absence of quotes, and note that of course whitespace is to be read as **y**, in quoted alien text. This proposal is not new for Linnaeans, in effect, but it is NEW as an implementation of strong quotation (the 1989 Loglan version is quite unparseable in BNF or PEG formats) and allows NEW constructions in the cases of **sue** and **sao** (**sao "ice cream"**, to be read **sao ice y cream**

(also transcribable in the latter form).

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

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

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

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

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

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

structure words: Structure words were to be recognizable because made up entirely of V, CV, and CVV units (Cvv-V being a later refinement). In any case, there would never be two successive consonants in a structure word, and structure words (and phrases made of structure words) would be for the most part flat evenly stressed sequences of these units. Vowel initial words presented special problems which required that one pause before them.

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

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

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

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

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

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

The next complication was the Great Morphological Revolution, with the attendant desire to build complex predicates out of units of the shapes CCV, CVC and also CVV – the last units presenting the problem that they do not enforce formation of a CC juncture. This caused the need for phonetic hyphenization of djifoa (which were originally called “affixes”), with **r** and **n** but also with **y** (the introduction of **y** doing some damage to the purity of our vowel system; I would like this vowel to have a more distinctive pronunciation, as John Cowan has already suggested). The construction of complexes is a complex enterprise, but the definition can be given and is

unambiguous. There is weirdness attendant on the doubled vowel rule, which restricts the positions in which certain djifoa can appear, and also of course the fact that a CVC djifoa cannot appear in final position.

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

Later revisions or clarifications are

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

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

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

orthography of names: Names with **la** are always to be written phonetically and cannot end with clusters of three consonants (this can usually be fixed because one of them is generally a continuant and can be doubled). Names are spelled differently than in the Sources rather often, but no names have been rejected. The interpretation of the old Linnæan name operator **lao** as handling foreign names made it possible to resolve the issue of whether names with **la** should be phonetic or preserve foreign orthography.

doubled vowels forbidden in borrowings: A minor point was that it proved convenient to forbid the doubled vowels with their attendant stress from occurring in borrowed predicates. If stress is not given explicitly, it has to be deduced from the orthography, and the doubled

vowels made this harder. Only one word had to be revised. Doubled vowels in complexes are annoying but we have a fair number of them in the vocabulary, so they just add eccentric charm.

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

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

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

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

Finally, I am very pleased to be able to present a genuine phonetic parser for Loglan as a mode of the ordinary parser. To achieve this, I had to make sure that one could write any Loglan utterance in a form with explicit stresses and syllable breaks where desired (non-final stresses being treated as a species of syllable break and so following the syllable rather than the

vowel) and with no whitespace except explicitly comma marked pauses. LIP did not actually support explicit commas in all cases where word breaks were mandatory! Because of the phonetic parser, we no longer need phonetic notation for the language (the one used in our Sources is a bit strange) and we have a native way to indicate stress which can be used in written text for emphasis. It is also charming that the rule about pausing after a finally stressed *mapua* before a predicate, which goes back to the beginning of the language, is actually expressible in the orthography!