



Logical Language Lojban

A Hackers' Spoken Language *

Sven Moritz Hallberg
Karl Hans Janke Kollaborativ
sm@khjk.org

November 2005

Abstract

Lojban is an artificial language meant to be spoken by humans. Its distinctive features include a basis in logic, regularity, an isomorphic mapping between written and spoken form, and unambiguous grammar. By providing a glance at Lojban structure and grammar, this paper aims to show that, just as we develop and use new programming languages for additional power, we can develop and use new languages for use among ourselves. While only the major concepts of the language are introduced in an overview style, pointers to the comprehensive learning material are provided.

What is Lojban?

la lojban. mo¹

Lojban is a constructed spoken language. That is, it was artificially created by humans to be spoken by humans. Its clear structure also makes it easy for computers to parse and produce. “Lojban” itself is a Lojban word derived from the words for “logical language”. In fact, the basic idea of Lojban is to form sentences and other “utterance” structures in the way of predicate logic, as in

is-going(*person, to-destination, via-path*) ,

*This is the paper version of a talk to be held at the 22nd. Chaos Communication Congress, 27–30 December 2005 in Berlin, Germany. See <http://www.ccc.de/congress/2005/>.

¹You can translate these subtitles with the excellent jbofi'e program (“Lojban-fish”). See [Fi'e].

or the more “Lojbanic” form

person is-going to-destination via-path .

To become a proper spoken language of full expressiveness, a number of constructs for combining, modifying, and nesting such predicate expressions are needed, the most important of which shall be introduced in due course.

As some will be aware, there exist many constructed languages, created for a variety of purposes. A prominent example from the realm of science fiction is Klingon. More seriously and probably most well-known today is Esperanto. Lojban is a rather young development, started in 1955 with its ancestor Loglan². The current version of the language has been finalized as late as 1997 with the publication of [Ref]. Here is a quick run-down of Lojban's major design features, as not already mentioned:

- Lojban is designed to be culturally neutral.
- Lojban grammar is unambiguous, yet flexible.
- Lojban has phonetic spelling, and unambiguously resolves its sounds into words.
- Lojban is simple compared to natural languages; it is easy to learn.

²Loglan was originally created by James Cooke Brown, published in *Scientific American* in 1960. Disapproved of by Dr. Brown, Lojban was “forked” from Loglan in the 1980's to support a broader and more open development of the language. Lojban uses different words and a slightly different alphabet than Loglan but is otherwise very similar to it.

- Lojban’s 1350 root words can be easily combined to form a vocabulary of millions.
- Lojban is regular; the rules of the language are without exceptions.
- Lojban attempts to remove restrictions on creative and clear thought and communication.

Why was Lojban created?

Lojban was initially developed as a tool for research on the so-called Sapir-Whorf hypothesis. Expressed in simple terms, this hypothesis states that the structure of one’s language constrains one’s thinking. Consequently, Lojban was designed to have the full expressive power of natural languages but a significantly different structure. Thus it can serve as a test vehicle for scientists to study the interactions of language, thought, and culture. But thanks to the nice features listed above, it is interesting much beyond that. To show this is one of the major...

Aims of this Paper

vi selci’a te gunka

My fundamental goal is to demonstrate that Lojban is of more interest than merely as an element of scientific research. An obvious argument is that there already exists a significant body of texts written in Lojban, some of which are translations and others entirely original. These examples encompass a diverse collection of fields, including poetry ([Poe]), narrative ([Sto], [Lit], [Alc]), and mathematical writing ([Alg]). Obviously, it is beyond the scope of this paper to discuss these applications in detail, but the interested reader is very much encouraged to consult the bibliography.

The subtitle of this paper reads “A Hackers’ Spoken language”. I shall now explain why: The very structure and design of Lojban display values quite well-known to us computer geeks³. It is clean, simple, general, and, by the very virtue of all these, powerful. The rest of the paper hopes to demonstrate this by introducing the most important aspects of Lojban grammar. Note that it is beyond scope to fully teach the concepts presented and some things will only be explained in passing.

³and mathematicians, mind you!

Please do not be confused by this, the paper is only meant as an introductory overview. If it sparks your interest you will find directions to the proper learning material in the appropriate section below.

In light of the above, I do hope to motivate the actual use of this astounding language by the audience. To this end, properties like Lojban’s robustness when spoken over noisy channels, its strong support for precise expression, and possible suitability as a human-computer interface language shall be stressed.

Outline

greku

After introducing alphabet and pronunciation used by Lojban in the next section, the basic form and building blocks of simple statements will be explained. After that the Lojban tense system and some other ways to make sentences more specific are introduced. Next in line are questions and logical connectives. A section containing pointers to learning material follows and the paper concludes with an outlook on possible advantages of Lojban in the context of computers.

Alphabet and Pronunciation

casnu le se lrfu .e le se bacru

The canonical⁴ Lojban alphabet is a subset of the Latin alphabet plus three punctuation characters:

’ , . a b c d e f g i j k l m n o p r s t u v x y z

As you can see, the latin letters ‘h’, ‘q’, and ‘w’ are not included. Most of the letters are pronounced similarly to German and possibly other European languages. English however, differs in more cases, especially the vowels.

Of the three punctuation characters, only ‘ ’ represents an articulated sound at all, like the english letter ‘h’ in the word “ahead”. The period marks a mandatory pause (which may be shortened to a glottal stop) between words and it can be left out in writing as long as the words are separated by a space. The comma is used to explicitly mark a syllable break in the written form of a word, usually

⁴Heheh.

'	[h]	ahead	m	[m]	bottom
.	[ʔ]	stop/pause	n	[n]	button
a	[a]	father	o	[o]	joke, note
b	[b]	bottle	p	[p]	powder
c	[ʃ]	shirt	r	[r]	letter
d	[d]	dance	s	[s]	soldier
e	[ɛ]	bet, lens	t	[t]	time
f	[f]	fall	u	[u]	boot
g	[g]	goose	v	[v]	voice
i	[i]	green	x	[x]	loch, Ach
j	[ʒ]	measure	y	[ə]	above
k	[k]	keen	z	[z]	magazine
l	[l]	bottle			

Table 1: Pronunciation of the Lojban alphabet

one that is not obvious to the reader. It must not be pronounced.

The pronunciation of the Lojban letters has been designed to make the individual sounds as distinguishable as possible. That is, they are as different from each other as possible. This gives tolerance towards background noise and deviation by the speaker from the recommended pronunciation. Table 1 lists the pronunciation of all characters using IPA notation along with english example words containing an approximation of the sound in question. It should be stressed that these are only the “recommended” sounds; there is considerable room for deviation. Especially the letter *r*, which many cultures pronounce quite differently, explicitly allows a rolled, trilled, or any other form of “rhotic” pronunciation. In general, you must only observe the rule that your pronunciation of each letter should not be mistakable for another.

Lojban phonology also contains diphthongs which arise naturally and are written just as the vowels that produce them. They are listed for completeness in Table 2.

Predicates and Arguments

casnu lo bridi .e lo sumti

As already mentioned in the introduction, Lojban’s sentence structure is based on predicate logic. A statement in Lojban consists of some predicate relationship which takes a number of arguments. The Lojban word for such predicates is *bridi* and the ar-

ai	[aj]	high	io	[jo]	Yolanda
au	[aw]	cow	iu	[ju]	beauty
ei	[ɛj]	bay	ua	[wa]	wander
oi	[oj]	boy	ue	[wɛ]	well
ia	[ja]	yard	ui	[wi]	week
ie	[jɛ]	yell	uo	[wo]	woe
ii	[ji]	hear ye	uu	[wu]	woo

Table 2: Pronunciation of diphthongs

guments are called *sumti*.

Our first example bridi will be *klama*, which means “to come to”. As example *sumti*, let’s use the characters *Pesco* and *Maya*. We will “lojbanize” their names, so they become proper grammatical Lojban words. Otherwise we would need to use escaping constructs⁵. To turn a name into its lojbanized form, we simply somehow translate the letters so the resulting pronunciation sounds like the original and, if the name does not already end in a consonant, we attach an arbitrary one of our choosing. The latter is needed for grammatical reasons; all Lojban names must end in a consonant. Also, they must always be followed by a stop, which we explicitly as usual by writing a period. Thus we obtain our example cast:

	Lojban name	sumti form
Maya	ma,iar.	la ma,iar.
Pesco	peskos.	la peskos.

Note that to turn the name into a *sumti*, the particle *la* is attached. This word has no translation; it is a grammatical marker meaning “name follows”. If it were absent, some confusion could arise. The same applies to the trailing period (pause). It avoids confusion about where the name ends.

Now we can build our first sentence which should mean “Pesco comes to Maya”. We know *bridi* and *sumti*, so how to arrange them? The basic form of a Lojban bridi statement is as follows.

x_1 cu *bridi* x_2 x_3 x_4 x_5

Obviously *bridi* represents the bridi. The particle *cu* is just a separator word which means “bridi follows”. It can be elided in most cases, but can be useful at times and it is never wrong. The x_i represent the *sumti* and their order is significant. Every

⁵which exist – useful if a name must be recognizable by spelling; think “Goethe”

mi	I (the speaker)
do	you (the one addressed)
ko	you-imperative
ti	this here
ta	that there
zo'e	something obvious or unimportant
ri	the previous sumti
di'u	the previous utterance

Table 3: Some common pro-sumti

bridi takes a certain number of them (but never more than five) and assigns its specific meaning to them. Number and meaning of the sumti places is called the bridi's *place structure*. For example, the complete place structure of klama is

x_1 comes to x_2 from x_3 via x_4 using x_5 .

So there, we just substitute our two sumti in the appropriate places (unused ones at the end are omitted) and obtain our first example sentence:

la peskos. cu klama la ma,iar.

“What if I want to fill x_3 but not x_2 ” you ask? Don't worry, you can explicitly specify the place a sumti should be associated with by prefixing it with one of the words fa, fe, fi, fo, and fu (for first through fifth place, respectively). For example, la peskos. cu klama fi la ma,iar. means that Pesco comes *from* Maya now.

There are yet other useful ways of modifying place structure, but I will not go into them now. Instead let us look at some more interesting ways of obtaining sumti. They employ more short “special words” like la, cu, fi, etc. These are called *cmavo*, or *structure words*. They bring and hold everything together.

Pro-sumti

Pro-sumti are analogous to pronouns in English. They stand for certain things and can be used in any sumti place. Table 3 lists the most important examples. Some of them need no explanation, but the other half warrants some words.

ko This word is used exactly like do but immediately turns the sentence into an imperative.

zo'e This placeholder can be used as a way of skipping over sumti places. For example, la peskos. cu klama zo'e la ma,iar. has the same meaning as the above version using fi.

ri, di'u The target of these “back-reference” words is usually clear, but very complex sentences can obscure what exactly is meant by “previous”. You shall be spared the details here, just rest assured that there is a clear rule.

Articles

This section introduces the Lojban *descriptor* word le, which is very close in meaning to the english definite article “the”. There are several more descriptors, for indefinite reference, referencing whole masses of things, etc.

Descriptors turn bridi relationships into sumti. The usual form is as follows:

le *bridi* ku

This phrase can be used in any sumti place and means some specific thing or things that would fit the x_1 place of the *bridi*, i.e. would make the relationship true if substituted for x_1 . The terminator ku is only for disambiguation purposes and can usually be elided.

Consider for example the bridi gunka with place structure

x_1 works on x_2 [activity] with goal x_3 .

The expression le gunka translates to “the worker(s)”. Note that le implies neither singular nor plural; there is a separate mechanism for specifying the number of objects meant.

If you want the descriptor to refer to another place than x_1 , use one of the cmavo se, te, ve, and xe. These have the simple effect of making x_1 switch places with one of the other sumti slots. For instance, se swaps the places of x_1 and x_2 . Thus le se gunka means “the work”.

Travel in Four Dimensions

nu klama fo lo se voncimde

Tense is usually understood to specify the location of some event in time, with respect to the speaker. In addition, Lojban employs the same concept to

pu	in the past
ca	in the present
ba	in the future
zi	short distance in time
za	medium distance in time
zu	long distance in time
<hr/>	
vi	short distance in space
va	medium distance in space
vu	long distance in space
zu'a	left
ri'u	right
ga'u	up
ni'a	down
ca'u	front
ne'i	within
be'a	north of
:	:

Table 4: Tense indicators

specify the spatial location of events. Tense (spatial or temporal) is attached to a bridi relationship by preceding the bridi with suitable cmavo. For example, **ba** means “in the future”, so **la peskos. ba klama la ma,iar.** means that Pesco will go to Maya some time after the sentence is spoken. The tense can be further modified by attaching a distance cmavo; **zu** for instance means a great temporal distance. Consequently **la peskos. ba zu gunka** means “Pesco will work in the remote future”.

Spatial tense works exactly like the temporal case, but includes cmavo for directions in three dimensions. Also, the distance cmavo are different. Table 4 provides an overview.

Tenses can be easily combined. You can attach a whole sequence of direction-distance cmavo pairs to your bridi. The meaning is to be understood as follows. Imagine a journey through space and time, starting at the present location of the speaker. Each cmavo pair gives one step of movement, in the order they appear. Wherever this sequence of steps takes you is where the relationship holds. Example:

la peskos. bazu zu'avi ga'uvu cu gunka

Pesco will in the far future work a little to the left and very far up from here.

Note that I have neglected to put spaces between

the direction and distance cmavo. The space between words is only for easy reading, Lojban grammar does not require it and it needs no pronunciation⁶.

Finally, the direction cmavo can also be left out, so **vi et al.** mean near etc. in space and likewise, **zi** and **friends** mean near etc. in time (past or future).

Be more specific!

le se cusku be ko cu zmadu de'u le ka steci

There are several ways of constraining the range of possible meanings of your utterances. I shall only describe some basic ones.

The earlier subsection on articles introduced *description sumti*, consisting of **le**, followed by some bridi. Recall that the resulting sumti then refers to things that could fit in that bridi's x_1 slot. In order to narrow down the set of such possible fits, the rest of the argument places can be filled with values. In order to avoid confusion, these *internal sumti* have to be marked, like so:

le bridi be sumti bei . . . bei sumti be'o ku

That is, the first internal sumti is marked with **be**, all later ones with **bei**, and the whole sequence is terminated by **be'o**. The terminator can again be elided if no ambiguity results.

Another important concept are compound bridi. Two bridi spoken directly after one another, say **klama gunka** form a new bridi. Such compounds are called *tanru* and their meaning is that of the second bridi, “modified” by the meaning of the first one. The exact nature of this modification cannot in general be specified and must be derived from the context. The place structure of a tanru is always that of the modified, i.e. last bridi. In the example above, **gunka** means “work” and **klama** means “to come” or “to go”. So the result must be some kind of work, somehow associated with going. It could be translated as the colloquial “legwork”. The place structure would be:

x_1 performs legwork x_2 for goal x_3

A significant hardship when learning a new language is how to express feelings and attitudes, because in most cases, these are merely communicated

⁶However, it is not permissible to split words with extra space! For instance, **gun ka** does not parse.

.ui	happiness	.uu	pity
.u'i	amusement	.u'u	repentence
.ue	surprise	.o'i	caution
.ua	discovery	.o'a	pride

Table 5: Some attitude indicators

as implicit connotations to other statements. To support direct expression of these, Lojban has a set of cmavo called *attitudinals*. The simplest way to use them is to put them before a sentence, thus specifying the attitude underlying that sentence. For example:

.ui la ma,iar klama
Yippie, Maya is coming.

Table 5 lists a number of common attitudinals.

Other specification constructs include several cmavo for associating two sumti, attaching entire bridi to sumti (i.e. relative clauses), quantifying sumti or bridi with cardinality or ordinality numbers, adding new sumti places to a relationship, etc.

From this it should be visible that the design of Lojban, while permitting very broad meanings, goes to great lengths to ensure that the speaker can also very easily specify and explicify as many aspects of his intended meaning as desired.

Questions

preti

Questions in Lojban are constructed quite easily. There are basically two kinds: Questions about the truth-value of some statement and “fill in the blank” questions. The former kind is built simply by prefixing an entire statement with xu:

xu la ma,iar. klama la peskos.
Does Maya come to Pesco?
(“Is it true that ...?”)

The more specific fill-in-the-blank questions are created by using special cmavo in the place of the “blank”. The question can ask for a sumti or bridi, using the blank words ma or mo, respectively.

la lojban. mo
What is Lojban?
(“In which relationships ...?”)

la peskos. klama gunka fi ma

What does Pesco do the legwork for?

It is of course also entirely possible to ask for multiple sumti at once:

ma klama ma
Who goes where?

Answering

Answers to fill-in-the-blank questions work just like in English. The sentence in question is repeated with the blanks filled in (and other now obvious parts possibly elided).

Lojban does not, however, have direct equivalents to the words “yes” and “no”. Instead, also truth questions are answered by repeating the statement in question, possibly negated. The negation particle is na and it is placed before the bridi. For example:

- xu la ma,iar. klama
 - la ma,iar. klama or la ma,iar. na klama

Obviously, repeating the whole statement word by word gets very tiresome. Luckily there exists a short-cut in the form of the cmavo go'i, meaning “the previous bridi”. Any sumti left out when go'i is spoken are assumed to be the same as in the original sentence. Thus go'i by itself has precisely the effect of answering “yes” to a xu-question and na go'i corresponds to “no”.

- xu la peskos. ba klama la ma,iar.
 - go'i or nago'i

Logic

logji

Lojban contains words for the full set of logical connectives. These come in different forms, depending on what they connect; sumti, bridi, sentences, etc. For the purpose of this paper, the basic ones for connecting sentences shall be shown. Oh, wait. A short discourse is in order, about how Lojban sentences are connected in general.

English separates sentences by punctuation. Lojban has no such punctuation marks and if, say, two

.i ja	\wedge	both true
.i je	\vee	either or both true
.i jo	\Leftrightarrow	both true or both false
.i jonai	XOR	exactly one true
.i naja	\Rightarrow	if first then also second true
.i janai	\Leftarrow	if second then also first true

Table 6: Logical sentence connectives

relationship statements were to be spoken one after another, confusion would probably result. So instead of punctuation, Lojban has a *word* for connecting sentences. It consists of only a single sound:

sentence .i sentence

The word *.i* has no logical connotation, although it can often be translated into english by an equally non-logical “and”.

In order to establish some logical connection between two sentences, the *.i* is followed by an indicator word. Table 6 lists the common cases. Try to find the pattern! The connectives for other kinds of utterances follow the same pattern, so everything is quite easy to memorize.

How to Learn Lojban

ma xe cilre fi la lojban

If you would like to learn Lojban, the usual material is either the so-called “Level-0 Booklet” [Lvl0] or the beginner’s lessons [Beg] for a more relaxed pace. To me, the Level-0 book felt somewhat “heavy” in the later parts about complex sumti and selbri constructions, but it is still definitely readable if one is a little impatient.

The basic grammar can be learnt within only a couple of days but of course you will need to build up some vocabulary. This mainly requires self-discipline and computer programs are available to aid you with it. At this stage, the most important thing is practice. You should try to find other Lojban newbies which is quite easy on IRC. Senior as well as junior members of the Lojban community regularly hang out on #lojban of the Freenode network. Also, the Logical Language Group will be happy to hook you up with other Lojbanists in your area. Just consult the Lojban website [Jbo] and send them an email.

For learning vocabulary, after ingesting most of the vocabulary in the Level-0 book, I have personally found it practical to attempt small translations or conversations, looking up required words and adding them to my personal word list. For this, the jbovlaste [Vla] dictionary interface is invaluable.

Once you have mastered the basics, the “Reference Grammar” [Ref] will become interesting. Despite its name, it should not be feared. Although it specifies all of Lojban, it is written as an accessible book containing valuable explanations and rationale for the finer points of the language.

Lojban and Computers

casnu la lojban. ne lo skami

This section provides some thoughts on possible uses of Lojban in different aspects of human-computer interaction. For all of these, obviously, some level of proficiency with Lojban must be assumed on part of the involved people. I have not deeply investigated any of these topics, but they might be worth some closer research.

Lojban grammar really is unambiguous. The authoritative reference is specified in YACC format⁷, and its unambiguity has been algorithmically confirmed. Because of this, Lojban input can easily be parsed completely for further processing. It would be quite possible to recognize questions and imperatives and implement a powerful natural language interface as a collection of handlers for patterns of such. More specialized input languages corresponding to appropriate subsets of Lojban would also prove useful.

It is very easy for a computer to produce correct Lojban. Even if the end user does not speak Lojban, it could still provide a flexible basis for an “i18n” framework. Current systems have problems with messages incorporating dynamically-generated values, because the structure of neither the base nor any of the target languages is understood by the program. Maybe using Lojban as the base language for translations would have benefits.

Lastly, Lojban could be used in programming, for naming functions, routines, etc. For logic programming languages like Prolog, the correspondence should be obvious. But also functions are

⁷A derived EBNF version is available.

- [Sto] *lo lojbo kamjikca lisri* (The Lojbanic Interactive Story)
http://www.lojban.org/texts/original/lojban_story/story/story.pl
- [Lit] Raymond Carver, translated by Jordan Delong: *Little Things* (1989)
<http://www.lojban.org/tiki/tiki-index.php?page=Little+Things>
- [Alc] Lewis Carroll, translated by the Lojban community: *Alice's Adventures In Wonderland*
<http://www.lojban.org/texts/translations/alice.html>
- [Alg] A. Adrian Albert, translated by Nick Nicholas: *Fundamental Concepts of Higher Algebra*, first page, University of Chicago Press (1956)
<http://www.lojban.org/files/texts/algebra>