

WORK IN PROGRESS

Systematically defined lojban vocabulary

Andrew Browne
Brisbane, Australia
dersaidin@dersaidin.net

2014-04-04

Abstract

*Lojban is a constructed logical language. Instead of nouns or verbs, root concepts are communicated in lojban using a vocabulary of about 1350 **gismu**. This paper discusses systematically analyzing the gismu to identify any issues such as inconsistencies and unclear definitions, and systematically designing changes to address issues identified. This paper does not specify what these processes should be, nor does it specify a policy for changes (such as backwards compatibility), but offers suggestions.*

Keywords

Lojban, Loglan, Logical Language

1 Introduction

1.1 Gismu

Instead of nouns or verbs, root concepts are communicated in lojban using a vocabulary of about 1350 **gismu**. The term gismu is used because Lojban gismu are used differently to nouns or verbs in English. Using terms such as nouns or verbs may be misleading and confusing.

Gismu specify a relationship among one or more entities. For example, here are three different relationships between the same two entities[1]:

- John is the father of Sam.
- John hits Sam.
- John is taller than Sam.

Lojban uses gismu to convey each of these concepts relating the entities, or **tanru** (built up from combinations of gismu) for more complex relationships.

Here are more examples with the same relationship ("give"), with the entities taking different roles in the relationship[1]:

- John gives Sam the book.
- Sam gives John the book.
- The book gives John Sam.

The meanings are different because the entities are in different places.

Lojban has a number of rules for generating the form (spelling/sound/word) a gismu may have. These rules prevent grammatical ambiguity, and assisted with generating the gismu set.

1.2 Problems with gismu

There are commonly several types of complaints about gismu in existing Lojban. [2]

- Unclear definitions
- Redundancy
- Orthogonal concepts mixed in one gismu
- Inconsistent place structures
- Not culture neutral
- Missing places intrinsic to the concept
- Missing a gismu for a concept
- Inconsistent choice between using **lujvo** for a concept or assigning it a gismu

What assurance do we have that the set of gismu are free of these issues?

1.2.1 Original gismu creation

The set of gismu was originally created using rules to systematically select a form ¹ for each gismu. These rules specified requirements on the spelling of the gismu, and selecting forms using sounds from the equivalent word in natural languages weighted by the number of speakers [3]. The mapping from a concept to a form is fundamentally arbitrary, and

¹The terms *form* and *spelling* are used interchangeably.

choices of form for any particular gismu are not considered to be a problem. These rules for forms are critical for lojban to be unambiguous.

In the gismu creation process, there were no rules² covering other aspects of gismu creation such as:

- Which concepts get a gismu (instead of a lujvo)
- Which concepts are related (share a gismu, overlapping places) or distinct
- Place structure

Not having rules for these other aspects enabled lojban to be created without having to overcome the significant effort of this additional analysis.

The concepts which gismu were assigned to, and their choice of place structures, were often based on the primitive words in Loglan (Lojban's predecessor), as well as words in natural languages [4]. Loglan also did not have a well specified systematic design, words were chosen arbitrarily[5].

Instead of a a systematic design, gismu were constructed arbitrarily. Although there was not a systematic design, the process was done with care by experts On the whole, so the resulting gismu list generally appears to be good quality. However, without a rigorous and well defined systematic approach, instances of the issues mentioned above slipped in.

1.2.2 Consequences

We claim that these problems undermine lojban's goals.

1.3 Towards better gismu

This document outlines a plan for improving the set of gismu. Having a systematic, well defined approach for analyzing and redesigning the gismu intends to be more systematic.

The planning is broken into four steps:

1. Establish goals for gismu (Section 2)
2. Establish process for systematic analysis of gismu against goals. find issues where goals are not met and changes are required (Section 3)
3. Establish policy for gismu changes (Section 4)
4. Establish process for changes to meet goals, in accordance with chosen policy (Section 5)

Once these goals and processes are established, the work on applying them to the gismu set can begin.

²As far as I can see, certainly none as well defined as the ones specifying form selection.

2 Goals for gismu

In order to evaluate gismu definitions we need establish a set of goals, or optimization criterion. The goals chosen for the gismu must be dictated by the goals of the lojban language, however these are not well defined.

I propose the following prioritized goals for an ideal gismu list (and by implication, for lojban more broadly):

1. Logical
2. Unambiguous
3. Clearly defined
4. Regular and Consistent
5. Atomic concepts
6. Complete
7. Minimal
8. Usable

2.1 Justifications for Goals

The rest of this section will elaborate on this choice of goals.

2.1.1 Logical

The name lojban literally translates to logical language. We claim being logical is generally accepted as the obvious primary goal of lojban.³

The goal of being logical implies the necessity of the other goals.

2.1.2 Unambiguous

Unambiguous parsing is a core feature of lojban's design.

2.1.3 Well defined

Words with no meaning associated with them are not useful.

2.1.4 Regular and Consistent

Natural languages are frequently irregular and inconsistent. The madness in this is evident to non-native speakers learning them.

2.1.5 Atomic concepts

To express an atomic concept clearly, the language should not require involvement of some orthogonal concept.

To express composite concepts, lojban provides tanru or lujvo. Building a tanru or lujvo of only intended concepts is difficult if concepts are not separated atomically.

³Or is being a language the primary goal? In "logical type of language", the root of the noun phrase is "language" not "logical".

2.1.6 Complete

Ideally everything should be able to be expressed. For that, the language needs to cover all concepts.

2.1.7 Minimal

”Simplicity is the ultimate sophistication”.

2.1.8 Usable

The language must be usable for humans and computers, in text or speech form.

2.2 Non-goals

Do not intend to change feel of language. Do not intend to change symbol allocation strategy (based on native languages) Do not intend to change to concept hierarchy (like *ithkuil*). Concepts fall under multiple categories, so that isn't perfect.

3 Procedure for analysis of gismu

These are examples. I doubt I've guessed the best set of questions we should actually use, and they are a little redundant here and there.

3.0.1 Justification for gismu to exist. What are the other options?

- Is it a primary/atomic concept? Is it tangled with some concept that should be orthogonal?
e.g. *fenki*: I'd say crazy/insane are the same, but are probably separate from frantic/frenzy.

- Is it the opposite to another gismu?
Why do we have separate gismu in some cases, and just use negators in other cases? e.g. *citno*, *nalci'o*, *tolci'o*; *gleki*, *badri*, *tolgleki*, *tolbadri?*; *fenki*, *tolracli*

- Why do we want/need this gismu for this over potentially equivalent *lujvo*?

If you had to use a *lujvo* for this concept, what would best *lujvo* be?

e.g. *linto*, *tilju*: IMO these should be *lujvo*. The base gismu should be something like "x1 has mass" (I'm not sure what existing gismu base would be. Not *junta* because weight is not mass. *grake*? Perhaps it is missing?). Then add *rasfi* of *barda* and *cmalu* (or some similar) to the base.

- When can we decide to make a gismu?

Can we say "this concept is not atomic, it is composed of these other two concepts, but we decide it is important enough to be a gismu anyway". Should the gismu then resemble the relevant *rasfi*? Can we say "lets have this gismu because it would be super convenient, despite some issues".

- Can you think of examples of two orthogonal aspects that you would put in the same place?

3.0.2 Justification of its place structure. Any extra/missing places for the core concept?

- Are all of the places essential to the core gismu concept? Or are some places really an orthogonal relationship?

The "under conditions" places are examples of being orthogonal to the core concept of the gismu.

- What plausibly important things were excluded? Is the core concept complete?

- Is the place structure consistent with other (similar) gismu?

- Can we add places which are not part of the core concept for convenience?

I'd say no, because then we could get stuff like the "under conditions" places.

- Are the most critical places first, and the most optional places last?

3.0.3 What distinguishes this gismu from similar/related gismu/lujvo?

This is after it should exist rather than the alternatives. Now we need to explicitly explain how it differs from each alternative or similar word.

- Are there any gismu that have potentially similar places? How do they differ?

e.g. *lo fasnu*, *lo se gasnu*, *lo rinka*, *lo se bredi*, etc.; hich of these is equivalent to "an event"? Why?

The definitions of these gismu should explain how these differ from each other.

e.g. *fenki*, *bebna*; can *bebna* x1 be an event?

e.g. *zgana*, *catlu*; how do these differ? the distinction is not clear from in the definitions.

- Can you think of an example where swapping this gismu for the potentially similar gismu/lujvo would cause the *bridi* to change from true to false.

3.0.4 Is the definition culture neutral?

- Does it refer to any culture specific icons or customs?
- Would it work for an alien culture?

e.g. *terdi*: is this Earth specifically? Would an alien race call their home planet *terdi*? I would expect a definition like "x1 is a terrestrial planet" (with maybe some other places like "...in system x2...", etc.). Earth would then be the human terrestrial planet, *remna terdi* or some *lujvo* of that. This would be similar to the way *lojban* names countries, *sralo gugde* = Australia.

4 Policy for changes to gismu

Once we have examined the gismu list, and identified issues, we need to choose policy for changes.

1. Compatible with existing *lojban*
 - Can deprecate gismu or gismu places
 - Can add new gismu
 - Might change gismu definitions for clarification
 - Might change gismu definitions to be more precise (excludes less precise meanings)
2. Logical
 - If this was top priority we would make breaking changes to *lojban* to fix logical issues.
3. Well specified
 - Vagueness is a key part of *lojban*, but critically it is explicit vagueness. New gismu definitions should be explicit about how vague they are.

Each change should be accompanied by a classification of how radical the proposed change is.

4.1 Consequences of change

4.2 Measures to smooth change

One disadvantage of change is that it will break the meanings of the existing corpus of *lojban*. There are some 800,000 sentences of *lojban* from years of (primarily) IRC discussions, mailing lists, and translations [6].

Because *lojban* is regular and relatively easily parsable, it may be possible to have a tool automatically update the text in the corpus to reflect changes. This idea has been used successfully for software [7].

However, unlike source code which must compile, *lojban* speakers may not consistently produce perfectly valid, well formed *lojban*. Parsing nearly-well-formed *lojban* would be an interesting challenge, and relevant to *lojban*'s goals of being computer friendly.

5 Procedure for changes to gismu

Symantic groups with same place structure?

5.1 What should a gismu definition contain?

Having established goals, what content in gismu definitions would best serve these goals.

Gismu definitions should include evidence that the gismu is logical. To do this I think the definition should include justification according to a systematic analysis (i.e. answer all these questions about the gismu).

If an alien civilization out there decides to make a logical language, how much of their gismu would look the same as *lojban*? If we're both following a systematic and logical process, if our concepts are atomic, I would hope to see many equivalent concepts with similar structures emerge (obviously physical/cultural differences would still cause plenty of differences). Would their gismu list be more logical? Could we justify our gismu list to them? What evidence could we provide that our gismu set is logical?

6 Conclusion

We are re-evaluating the gismu from 10 years ago. We are doing this because we believe there are issues with the gismu list produced 10 years ago. 10 years from now, what will people think of the list produced now?

Why do we have the gismu we do? Why do we not have the gismu we don't have? In 10 years, will our revised gismu set hold up to the same questions?

We need processes to create a resulting gismu list that is logical. We need to be able to show evidence that the resulting gismu set is logical, to convince others that it is logical. We need clear and comprehensive definitions, so it can be understood more easily. If we are not systematic about this process, the revised gismu set may need to be revisited like this in another 10 years. We can do better than that.

References

- [1] "A quick tour of *lojban* grammar, with diagrams." http://dag.github.io/c11/2/1/A_Quick_Tour_of_Lojban_Grammar_With_Diagrams, 2012.
- [2] "Bpfb section: gismu issues." <http://lojban.org/tiki/tiki-index.php?page=BPFK+Section:+gismu+Issues> BPFK Section: gismu Issues, 2014.
- [3] "The gismu creation algorithm." <http://dag.github.io/c11/4/4/gismu>, 2012.

- [4] “Lojban mailing list.” <https://groups.google.com/d/msg/lojban/h6yQDGV51Qw/mE7K4vdMoZgJ> Lojban Mailing List, 2014.
- [5] *LOGLAN 1: A LOGICAL LANGUAGE 4th Edition*. 1999.
- [6] “korpora zei sisku.” https://www.alexburka.com/~danr/korpora_zei_sisku, 2014.
- [7] “Large-scale automated refactoring using clangmr.” <http://research.google.com/pubs/archive/41342.pdf> korpora zei sisku, 2011.