Predicate logic is a suboptimal basis for real time logical language

Abstract

The constructed language Lojban is based on modern theories of human communication. It was hypothesised in a previous issue of this journal that if it fails to become a viable tool for communication, natural languages – being suitable – must have properties we have not yet noticed.

This paper examines Lojban's predicate logic grammar to look for an explanation for its learnability problem. Previous suggestions include that the requirement to make unambiguous sentences is excessively demanding for the human mind. We refute this argument by formulating an ad hoc grammar based on natural language. This reductionist approach offers a logical model that is simpler than one based on predicate logic.

The evidence suggests that predicate logic is a suboptimal model for a logical language designed for real time use. A more adequate model can be extracted from natural language grammar.

1. Introduction

When Geoffrey Sampson (1999) reviewed John Woldemar Cowan's *The Complete Lojban Language*, he advised linguists not to treat the topic as a matter of mere curiosity. Sampson was keen to observe whether the circle of enthusiasts of the artificial language would prove capable of turning Lojban into a functional communicative medium among them.

Lojban was designed to be a language that makes predicate logic speakable, but seems complicated and strange compared to natural languages. This leads to a certain irony in that Lojban is based on important theories about human communication. (Ibid.)

As to the question whether Lojban would prove to be a fully functional communicative medium, Sampson foresees two alternative outcomes:

A: If Lojban proves to be a viable tool for communication, the question will arise why natural languages are not more like Lojban?

B: If Lojban fails to become a viable tool for communication, what differences between Lojban and natural languages make the latter but not the former usable? If the language fails, he suggests, natural languages must have other crucial properties we have not yet noticed. (Ibid.)

We will evaluate Lojban's fitness for a functional communicative tool (3). The evidence suggests that while some individuals have succeeded in keeping up real time conversation in Lojban, the vast majority of the Lojban community have failed to gain fluency.

We consequently accept a weaker form of Sampson's proposition A, by which we mean that in the following sections we will look for an answer to the question what characteristics in Lojban's grammar make it unusable *for the average person*.

Explanations for the learnability problem have been offered prior to this paper. Goertzel (2013) suggests that many individuals attempting to learn and use Lojban have found that communication is slow and difficult due to the limited body of existing vocabulary. In order to discuss Goertzel's suggestion we will examine grammatical complications within Lojban's word lists (3.2).

Okrent (2009) suggests that the very requirement of having to make logically unambiguous sentences makes using Lojban excessively demanding for the learner and the experienced speaker alike. She provides examples of the problems Lojban users face in everyday situations as well as in academic writing.

However our point of view is that some natural languages and schematic *a posteriori* languages have a relatively low level of syntactic ambiguity. Our hypothesis is that, as

opposed to Okrent's suggestion, Lojban's learnability issue is not caused by having to speak logically per se, but more specifically by having to do it with a grammar based on predicate logic.

To bring evidence for our claim we will construct an alternative ad hoc model (5.2) that is syntactically unambiguous and based on natural language grammar. With this model we will aim to prove that the complications caused by using predicate logic as a basis are avoidable.

2. LOJBAN'S THEORETICAL BACKGROUND

Artificial languages can be divided into two main groups: *a posteriori* and *a priori*. The grammar of an a posteriori language, such as Esperanto, Ido and Interlingua, is based on one or more natural languages. In contrast, a priori languages, such as Lojban, are experimental languages that can be based on the rules of logic, or such that aim to reflect reality more clearly than natural languages (Libert 2000).

Despite recurring advice to seize the potential of *a priori* languages for the purposes of theoretical linguistics, as to date, little research has been done. This could however be due to the learnability problem rather than a mere lack of interest.

Holeš (2004) argues: *"If a universal property holds even for artificial a priori languages, we can assume that this may be a property of any human language. These extraordinary products of the human mind, resulting sometimes from a deep analysis of the surrounding world, should be of interest to experts on general linguistics, language acquisition, and cognitive science."*

The logical language project was started to provide a tool to test the Sapir-Whorf hypothesis according to which languages constrain the world views of their speakers. For this purpose the logical language Loglan, predecessor of Lojban, was made "hyper logical". The standard criticism logicians make is that natural languages are too unreliable to allow reasoning without risk of being led astray by language structures. The developers' conjecture was that

a speakable logical language could provide a more suitable platform for logical thinking. (Smith 1991)

More recently Lojban has been proposed as a tool for artificial intelligence research because sentences made in Lojban are syntactically unambiguous and readily reducible to predicate logic (Goertzel, Pennachin & Geisweiller 2014).

As in predicate logic, statements in Lojban are made with predicates and arguments, but the range of expression has been expanded far beyond truth statements. Lojban does not have the standard parts of speech. Instead, predicate words can serve as the equivalent of a noun, verb, adjective or adverb. There are additionally a variety of connectives and other types of operators, including the kinds used in formal logic. (Cowan 1997)

Historically Lojban is a successor of Loglan, the first attempt to create a logical language. Sociologist James Cooke Brown started developing Loglan in 1955, and others joined the project prompted by an article in *Scientific American Magazine* in 1960. At that time a revolution of linguistics was taking place. (Ibid.)

This revolution was one that brought mathematical methods into the theory of language. A work that proved most influential to the logical language project was Quine's *Word and Object*, 1960. Other influences include the works of Zellig Harris, Victor Yngve and Noam Chomsky. (Brown 2008)

The spirit of the age is well crystallised in a quote from Montague Grammar: *“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory”* (Montague 1970).

3. Evaluating Lojban's success

Well over a decade after Sampson's (1999) review, it is hard to find an unequivocal answer to his question whether Lojban has proved to be a functional tool for communication.

Lojban has been developed for decades by dozens of workers and hundreds of supporters, led since 1987 by the Logical Language Group which estimates the number of enthusiasts at around two thousand people and boasts an active mailing list and an IRC chat room for text-based communication. (Logical Language Group 2014)

Conversely the number of fluent speakers has been estimated to be low. According to Goertzel (2013), several dozen users are fluent in writing, but unofficial estimates of the number of fluent speakers have been set at around half a dozen or lower (Eppcott 2008). This means in practice that the vast majority of the people who have been involved in creating the language have failed to gain fluency.

One concern raised within the community is that it has become customary to use a dictionary while speaking Lojban. The modest language skills of Lojban speakers have been contrasted to the number of people successfully gaining fluency in foreign languages as well as in other constructed languages, such as Esperanto. (Arnold 2006)

It has been suggested that the difficulty of learning Lojban is a direct obstacle to its original idea of being used to test the Sapir-Whorf hypothesis. People drawn to Lojban are usually predisposed to logical thinking, often engineers and mathematicians, and nevertheless have trouble learning the language. Teaching it to normal people in the purpose of making clinical tests seems unrealistic. (Okrent 2013)

Conscious efforts to focus on fluency have been a trade off on the expense of logical expression. Speakers are reported to have a tendency to simplify the grammar in live conversation to a point where the syntactic structure shares more similarity to natural languages than to predicate logic (cf. Powell 2010).

This could mean that establishing a living logical language community is impracticable. Although proponents claim that there has been some successful use of Lojban as a tool for real time communication, there is at least a lack of evidence suggesting that Lojban learning is taking off at a pace that can be expected from a language designed for human use.

We consequently propose a weaker formulation of Sampson's second conjecture (1). In this paper we will look for an answer to the question why Lojban, unlike natural and a posteriori languages, is not usable for the average person.

3.1 Method

We will first analyse the complexity of Lojban's grammar (3.2). We will then formulate an alternative a posteriori grammar that can function as the framework of a syntactically unambiguous language (5). In the following chapter (6) we will compare these two languages.

3.2 Complexity of Lojban's grammar

There is a formal Lojban grammar available as a computer file which consists of approximately 1,200 command lines (Cowan 1997). The dictionary gives many additional rules about word usage. Lojban has two open-end word classes: predicates and lexicalised proper names, and an additional 120 classes of particles (Sampson 1999). The total number of particles is 1,091. These include pronouns, connectives, prepositions, quantifiers and many other types of varying importance for syntactic functionality (see Logical Language Group 2002).

A matter to observe in detail is the predicate valency. Each lexical predicate word has between one and five argument places with various semantics determined word by word in the dictionary. These argument places should be well memorised to allow fluent conversation.

There are 3,388 assigned argument places altogether in a list of 1,432 basic predicate words. The most common type is one with two arguments. For example the dictionary entry *badna* x1 is a banana/plantain (fruit/plant) of species/breed x2. The first argument place (x1) is located before the predicate, similarly to the English subject. Other argument places follow the predicate. (Logical Language Group 1994)

An example of a predicate word with all five argument places is *draci* (a theatre play) which has the following arguments: a play (x1), a plot of play (x2), an author of a play (x3), an audience of a play (x4), an actor of a play (x5). (Ibid.)

To illustrate how the valency works, an equivalent of a sentence with the predicate *draci* could be expressed with the following English words: *event - PLAY - Hamlet - Shakespeare - The Globe - Jeremy Irons*, with the predicate word capitalised here for illustrative purposes. In other words: "Jeremy Irons is playing Shakespeare's Hamlet at The Globe". The logic of predicate grammar does however allow different translations. Another possible wording for the statement is: "Shakespeare's drama Hamlet is being played at The Globe by the actor Jeremy Irons".

But what if we only want to say that Jeremy Irons is acting? This can be solved by saying *something - PLAY - something - something - something - Jeremy Irons*.

To avoid tautology the arguments do not have to appear in the default order. Particles can be used to indicate an unexpected argument place, and unassigned arguments towards the end of the phrase may be omitted. The particle *fa* may be used to indicate that the next argument place is x1, and the particles *fe, fi, fo* and *fu* to indicate x2, x3, x4 and x5, respectively. (Cowan 1997)

A second option is to use the particle *se, te, ve* or *xe* before the predicate to indicate that the argument preceding the predicate word is x2, x3, x4 or x5, respectively, not x1 as is the case by default. The x1 argument place is pushed into the niche that becomes vacant, that is x2, x3, x4 or x5. (Ibid.)

As handy as these solutions may seem on paper, it can be very difficult for a learner, still struggling to remember the 3,388 default argument places in the predicate word list, to capture the meaning of sentences that may use the argument series defined in the dictionary as well as any of the aforementioned particles to make any changes to that order.

To illustrate the difficulty of the task, there are dozens of different types of argument places that may express time, place, agent, object, material, medium, method, result as well as many others. The first argument typically represents the semantics of the root word, but if we assume a modest twenty possible argument types for each of the following places, there are 20^4 or 160 000 different ways to arrange the argument series of just one predicate word which sums up to 160 million possibilities per 1000 words (cf. Logical Language Group 1994).

Predicate words could be categorised into subclasses according to their argument series, but this has not been done to date, and it would seem likely to produce a great number of very small groups because such a grouping has not been part of the language design. There is no set limit to the number of such subclasses, and a larger dictionary could contain further argument types.

4. PREVIOUS HYPOTHESES

Goertzel (2013) suggests that expanding the dictionary could facilitate language acquisition because limitations in the existing body of vocabulary makes using Lojban slow and difficult. Goertzel proposes using English as a source of predicate words. This way the number of arguments per predicate could be limited to a maximum of three, as in English ditransitive verbs[1].

According to Goertzel (2006) conventions for argument places will have to be created. He provides a sketch of 26 possible subcategories grouped according to their valency.

[1] Goertzel (2013) gives the English verbs *give* and *take* as examples of ditransitives.

However, increasing the number of dictionary entries would counteract the gain attained from limiting valency to a maximum of three arguments. Another question is whether such a revision would compromise Lojban's fundamental principles.

A different hypothesis is provided by Okrent (2009) who suggests that Lojban is made excessively difficult by the very requirement of having to form logically unambiguous sentences. The syntax is "exhaustively" unambiguous, because in order to choose the right wording one must clearly specify the structure as a whole, using markers that serve as spoken parentheses.

Okrent compares composing a sentence is Lojban to writing a line of computer code: choosing the wrong function, dropping a variable, or forgetting to close a parenthesis leads to a failure. (Ibid.)

5. An alternative approach: reductionism

To test Okrent's hypothesis we will construct a syntactically unambiguous a posteriori grammar. It is essentially a matter of arranging nouns, verbs, adjectives and other parts of speech known to exist in many natural languages in an ideal way. With such a grammar it will be possible to compare Lojban with a naturalistic approach hoping to answer the following questions:

(i) Does a syntactically unambiguous language have to be based on predicate logic?

(ii) Would an unambiguous grammar based on natural language have to be more complicated?

(iii) Can the complication of having word-specific definitions of argument places be avoided with a naturalistic approach?

Human languages are known to be irregular, complicated and ambiguous, at least to a degree. This is why we are looking for a grammar that is the highly regular simplification of

human language; a language with a minimal number of grammatical rules with no exceptions to facilitate controlling ambiguity.

For instance, an ordinary English phrase such as *I love you* is by no means syntactically ambiguous, provided that we postulate the rules as such that the verb is always preceded by the subject, and always followed by the object. Furthermore, we will need a way to identify *love* as a verb.

In this paper we propose a reductionist approach which allows one to work all the way from a very complicated natural grammar to a very simple model. We call this model Reduced Natural Grammar (RNG).

To make the task simpler we will use Esperanto as a starting point instead of a non-constructed language, such as English. As an a posteriori language, Esperanto is based on natural languages, most notably Latin, French, German, Russian, English and Greek (Kiselman 2008). As a highly schematic language, Esperanto, the creation of Dr Zamenhof (born 1859, died 1917), already has a grammar that is reduced significantly from the source grammars.

One particular insight by Zamenhof are the part of speech endings, including the noun ending *-o*, coined in the fashion of the Latin *-us* ending (*domus*, 'a house') which is shortened to *-o* in Italian (*duomo* 'cathedral'). Some additional endings in Esperanto are *-a* for adjective, *-i* for verb infinitive and *-e* for adverb (Kiselman 2008).

Different ways to indicate part of speech are found in natural languages, but when implemented systematically, it simplifies the task of parsing dramatically. Unlike Esperanto, RNG may assign a different ending for each and every part of speech.

There can be many possible ways to create an unambiguous grammar using a reductionist approach. Perhaps not the simplest, but we have chosen to construct a grammar on the SVO (subject – verb – object) word order. We believe this is a good choice in the topic of

logical languages because it essentially agrees with the basic word order convention of mathematical equations such as 1 + 2 = 3.

5.1 Defining relations between parts of speech

In 3.2 we have demonstrated that the majority of Lojban's grammatical rules are in the valency indications found in the vocabulary. The ideal way to avoid a similar complication for RNG is to create a universal dictionary with no valency indication. Such a dictionary will consist of the mere semantic values of idea words. Words are made syntactically usable by the systematic process of applying a different ending for each part of speech that is needed to render all kinds of utterances.

We have chosen noun as the default part of speech from which all other types will be derived. To create an ad hoc vocabulary we will use Esperanto root words with any necessary additions or modifications. We have chosen the x-convention for the orthography which means using the letter x to replace circumflexes (eg. *sxorto* for the Esperanto word *ŝorto*, 'shorts').

Six rules are needed; D stands for derivation rule:

D1. The default morpheme, marked with the noun marker (-o ending), expresses its corresponding idea; eg. *domo* 'a house', *jxeto* 'a throw', *belo* 'beauty'.

D2. Each morpheme marked with the adjective marker (-a ending) expresses having the property of the idea; eg. *bela* 'beautiful', *doma* 'houselike'.

D3. Each morpheme marked with the verb marker (-i ending) expresses doing what the idea does; *jxeti* 'to throw', *domi* 'to house', *beli* 'to impress beautifully'[2].

D4. Each morpheme marked with the adverb marker (locative -e ending) expresses loosely being within the idea; *hejme* 'at home', *morgauxe* 'tomorrow', *multe* 'much; in a big quantity', *nee* 'not'.

[2] The causative can be expressed with the suffix *-igxi,* cf. table 1.

D5. Each morpheme marked with the preposition marker (apostrophe) expresses being located within the idea which belongs to X (i.e. the following noun phrase); *sur' domo* 'on top of a house' (cf. *suro,* 'top'), *post' morgauxo* 'after tomorrow' (cf. *posto*, 'what comes after'). The description can also be formulated using the possessive preposition *de'* 'of': *sur' domo* = *sure de' domo* 'atop of a house'*; post' morgauxo* = *poste de' morgauxo* 'in the after of tomorrow'.

D6. Morpheme clusters composed with the compound binder (hyphen) agree with the following rule: each clustered morpheme modifies the one after it; *briko-domo* 'brick house', *sablo-sxtono-muro* 'sand stone wall'. This rule can be used to produce all other necessary types of nouns, adjectives, verbs, adverbs and prepositions needed (see table below).

type	**cluster**	**explanation**	**English equivalent**
noun	*domo-aro*	'house-set'	houses
	jxeti-anto	'subject of throw'	thrower
	jxeti-ato	'object of throw'	thrown (on purpose)
	jxeti-igxo	'undergoer of throw'	thrown (incidentally)
verb	*jxeti-igxi*	'act as the undergoer of throw'	be thrown
	jxeti-igi	'throw-cause'	make (someone) throw
adjective	*nacio-aro-intera*	'nation-set-inter-ADJ'	international
	salo-gusta	'salt-taste-ADJ'	salty
	nutro-hava	'nutrition-have-ADJ'	nutritious
	unuo-esma[3]	'one-ordinal place-ADJ'	first
adverb	*rapida-maniere*	'quick-manner-LOC'	quickly
	tago-mezo-antauxe	'day-mid-before-LOC'	AM (before midday)
	kio-tempe	'what-time-LOC'	when?

[3] The ordinal suffix *-esma* is borrowed from Ido grammar; see http://idolinguo.org.uk/bgrammar.htm.

	kio-ece	'which-binary state-LOC'	whether..?
preposition	*al'-en'*	'to-in'	into

Table 1: examples of types of compounds produced according to rule D6. For conjunctions, see rule S10 (5.2).

5.2 RNG syntax

Syntactic rules with no exceptions are needed to eliminate ambiguity; this is in practice the matter of allowing only one way to parse each sentence. RNG sentences are essentially arrangements of the parts of speech defined in 5.1. As we have chosen to construct an SVO language, we will use the following set of syntactic rules; S stands for syntactic rule:

S1. The first verb (D3) of each clause is marked by the predicate marker (*-as* ending, which replaces the infinitive *-i* ending).

S2. A noun phrase in the nominative, placed before the first verb, is the subject of the clause; ***hundo*** *kuras* 'a dog is running'.

S3. A non-verbal word phrase in the nominative which is placed after the verb is a subject complement; *hundo estas* ***granda*** 'a dog is big'.

S4. When there are two or more verbs in a clause, the latter verb modifies the previous verb; *hundo* ***volas provi kuri*** 'a dog wants to try to run'.

S5. Two consecutive nominal noun phrases (D1) are in apposition; ***sinjoro profesoro*** 'Mister Professor', ***tio hundo*** (that dog; the order can be inverted with no change in meaning).

S6. Each adjective (D2) refers to the whole following noun phrase (i.e. until the noun) when applicable; *tio* ***bela*** *granda hundo* 'that beautiful big dog'.

S7. Each preposition (D5) refers to the whole following prepositional phrase (i.e. until the noun); ***kun' tio granda hundo*** 'with that big dog'.

S8. Each adverb (D4) refers to the very next word; *kun' tio* ***tree granda*** *hundo* 'with that very big dog'.

S9. Each adverbial (cf. prepositional phrase) refers to the previous non-adverbial constituent; *tio hundo volas* ***kuri en' arbaro*** 'that dog wants to run in a forest'.

S10. The preposition *ke'* refers to the next predicate verb as a complement to the previous predicate verb (a conjunction, cf. D5; overruling S7); *Nio* ***venas*** *al' hejmo* ***post'-ke'*** *nio kuras en' arbaro* 'We will come home after we run in a forest'.

S11. The preposition *ku'* refers to the next predicate verb as a complement to the previous predicate verb before *ke'*-phrases placed between them (cf. S10); *Mio* ***sidas*** *en' interno se-ke' pluvas* ***cxar-ku'*** *mio volas* 'I sit in if it rains because I want [to sit in].

Time and mood can be expressed with adverbs and adverbials when needed.

Together these eleven syntactic rules suffice to allow the expression of complex unambiguous sentences, for example:

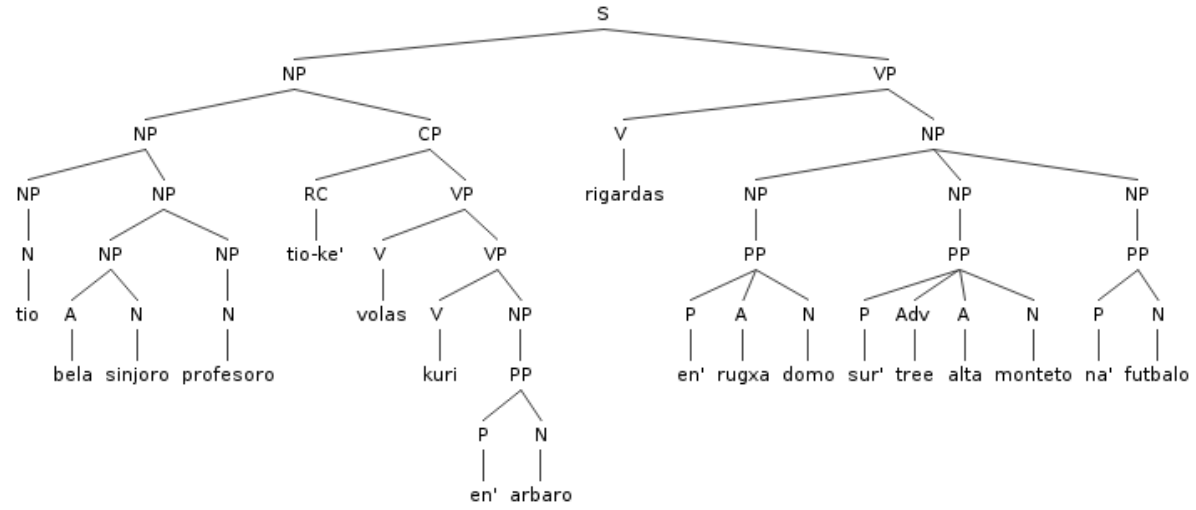

Diagram 1: The word derivation rules D1-D6 and syntactic rules S1-S10 allow only one way to parse each sentence.

tio	bela		sinjoro	profesoro		tio-ke'		volas		kuri	
this	beautiful		mister	professor		that-CONJ		want-PRED		run	
en'	arbaro	rigardas		en'	rugxa	domo	sur'	tree	alta	monteto	
in	forest	watch-PRED		in	red	house	on	very	high	hill	
na'		futbalo									
OBJ-PREP		football									

'This beautiful Mister Professor who wants to run in a forest watches football in a red house on a high hill.'

Recursion of adverbials is prohibited by rule S9, but it occurs elsewhere. For instance, D2 allows the derivation of adjectives from prepositions and adverbs, and S6 allows adjectives to modify other adjectives (see also S6 for compounding):

Petro-dea	filino-dea	domo-dea	pordo-dea	sxlosilo
Petro-of-ADJ	daughter-of-ADJ	house-of-ADJ	door-of-ADJ	key

'The key to the door of the house of Peter's daughter.'

Relative clauses can be used to allow the kind of recursion that may occur in English adverbials[4]:

mio	vidas	na'	domo	tio-ke'	estas	cxe'	lago	tio-ke'
I	see	OBJ-PREP	house	that-CONJ	is	at	lake	that-CONJ
estas	sur'	monteto	tio-ke'	estas	malantaux'	arbaro		
is	on	hill	that-CONJ	is	behind	forest		

'I see a house which is by a lake which is on a hill which is behind a forest.'

Recursion takes also place in object clauses.

mio	kredas	na'-ke'	vio	komprenas	na'-ke'	mio	nee
I	believe	OBJ-CONJ	you	understand	OBJ-CONJ	I	not
volas	na'-ke'	vio	iras				
want	OBJ-CONJ	you	go				

'I believe you understand I do not want you to go.'

Here the conjunction *na'-ke'* 'that' incorporates the object preposition *na'* as a prefix.

6. Comparison of Lojban and RNG

We will now compare structures produced by Lojban and RNG to find answers to questions presented in section 5.

A key difference between Lojban and RNG is in the way arguments are expressed. While the Lojban dictionary determines the number and types of argument places for each

[4] A variety of demonstrative pronouns (cf. D1) are needed for disambiguation. This paper, however, uses only the words *cxitio* 'this' and *tio* 'that'. The latter always refers to the last mentioned noun.

predicate, RNG has no such rules, and instead uses prepositions to form adverbials which give information about the verb.

Both Lojban and RNG place the subject before the predicate word or predicate verb, respectively, but in Lojban the subject can be moved to a different position using a variety of particles (3.2). In the RNG model designed to produce an SVO syntax (see 5), ambiguity is avoided by the more straight-forward rule of not allowing the subject to appear after the verb.

To give an example, the Lojban predicate *vecnu* 'sell' has four out of five possible argument places: x1 (seller) sells x2 (goods) to x3 (buyer) for x4 (price). The x1 argument place is the equivalent of the subject in English and is placed before the predicate word, the equivalent of the verb. Other arguments, placed after the predicate, share resemblance with the direct and indirect object of English; (Logical Language Group 1992:)

mi	cu	vecnu	**ti**	**ta**	**zo'e**	vau
I	SEPARATOR	sell	this	that	something	TERMINATOR

'I sell this to that buyer for some price.'

The separator *cu* and terminator *vau* are provided to stop compounding of two consecutive predicate words. In this example, with the sole predicate word *vecnu*, there is no such risk of confusion, so they can both be omitted. Also, the unspecified argument *zo'e* 'something', placed at the end of the predicate phrase, can be omitted. The sentence can thus be expressed in a simpler manner (Logical Language Group 1992):

mi	vecnu	**ti**	**ta**
I	sell	this	that

'I sell this to that buyer.'

The RNG model uses a different strategy:

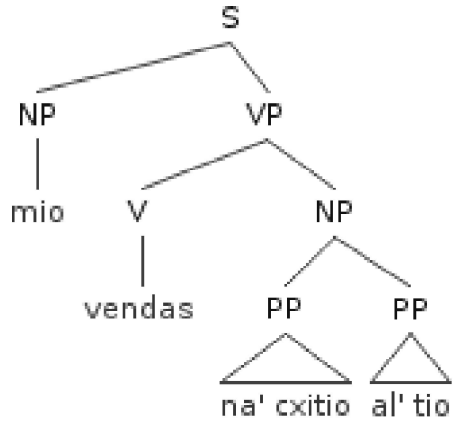

Diagram 2: the object noun phrases referring to the verb according to RNG's syntactic rule S9 (5.2).

mio	vendas	na'	cxitio	al'	tio
I	sell	OBJ-PREP[5]	this	to	that

'I sell this to that.'

In RNG, instead of having a set valency, the verb may be followed by any number of prepositional phrases which may appear in any order (see rule S9 in 5.2):

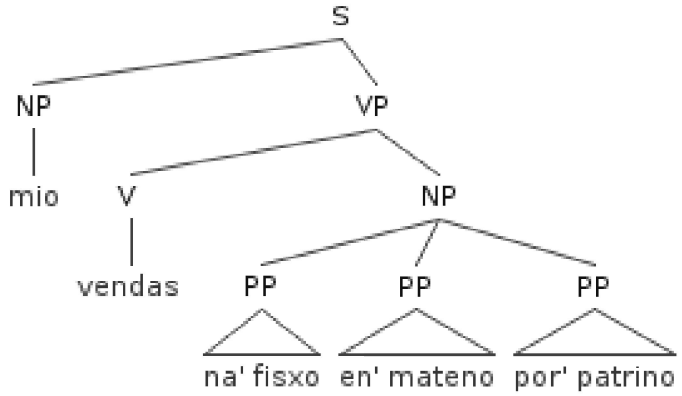

Diagram 3: increasing the number of arguments makes a minimal impact on the syntactic complexity of RNG.

mio	vendas	**na'**	fisxo	**en'**	mateno	**en'**	bazaro
I	sell	OBJ-PREP	fish	in	morning	in	market
por'	patrino						
for	mother						

'I sell fish in the morning at the market for mother.'

[5] The unofficial preposition *na* is discussed in Wennergren 2007.

The Lojban example *vecnu* however has no argument place for time, place or employer. Whenever necessary, prepositions are used to add a non-default argument.

The following example has the predicate *citka* 'eat' which only has two argument places: x1 eats x2. The particle *vi* 'short distance' may be used to express place (Cowan 1997):

le	ratcu	cu	**citka**	le	cirla	**vi**	le	panka
the	rat	SEPARATOR	eats	the	cheese	SHORT DISTANCE	the	park

'The rat eats the cheese near the park.'

The above example demonstrates that Lojban does not rely completely on the syntax of classical predicate logic. In addition to the dictionary-assigned valency, there are a further 65 prepositions in the particle list (see Logical Language Group 2002), which are used whenever the default arguments fail to convey desired meaning. Without these prepositions there would have to be a huge number of set arguments for each predicate word.

Lojban's structures share similarity with English verb valency. While English verbs control one to three set arguments, the corresponding number in Lojban is one to five; other cases are expressed with prepositional phrases in both languages. The Lojbanic system is however much more difficult to master because the argument semantics contain much more variety than just direct and indirect objects, as has been demonstrated above.

Going back to the previous example, we observe a second complication in the grammar (Cowan 1997):

le	ratcu	**cu**	citka	le	cirla	vi	le	panka
the	rat	SEPARATOR	eats	the	cheese	SHORT DISTANCE	the	park

'The rat eats the cheese near the park.'

This time the separator *cu* is necessary to avoid ambiguity that would otherwise be caused by two consecutive predicate words. In total the grammar contains 24 separators and terminators of various type and importance. They are sometimes compulsory and sometimes omitted – marked in square brackets in the following example (Logical Language Group 1992):

le	tavla	be		la .an.	bei
the	talker	INTERNAL ARGUMENT LINK		Ann	LINK MORE ARGUMENTS
le vecnu		[ku]		[be'o]	
seller		[END DESCRIPTION ARGUMENT]		[END INTERNAL ARGUMENT]	
[ku]			cu		klama
[END DESCRIPTION ARGUMENT]			ARGUMENT SEPARATOR		go
[vau]					
[END OF ARGUMENT LIST]					

'The talker to Ann about the seller goes.'

Learning to use such particles in live conversation demands good analytical thinking skills. But it is in the interest of this study to prove that such grammatical complications can be avoided in unambiguous language. The last two examples can be expressed in a simpler way in RNG:

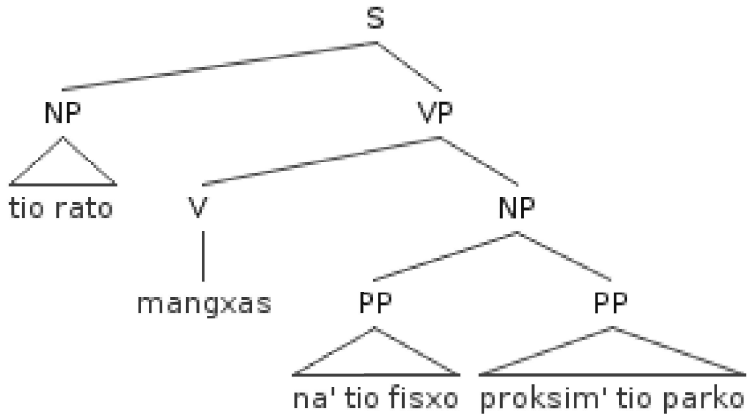

Diagram 4: RNG uses grammatical endings to identify a noun phrase (*-o*) and a predicate verb (*-as*).

tio	**rato**	**mangxas**	na'	tio	fromagxo	proksim'	tio
that	rat	eats	OBJ-PREP	that	cheese	near	that
parko							
park							

'The rat eats the cheese near the park.'

The words *rato* and *mangxas* are distinguished by their endings: *-o* for noun and *-as* for predicate verb; no further separators are needed. Furthermore, the multiple parenthetical particles of the second Lojban example are avoided completely in NRG:

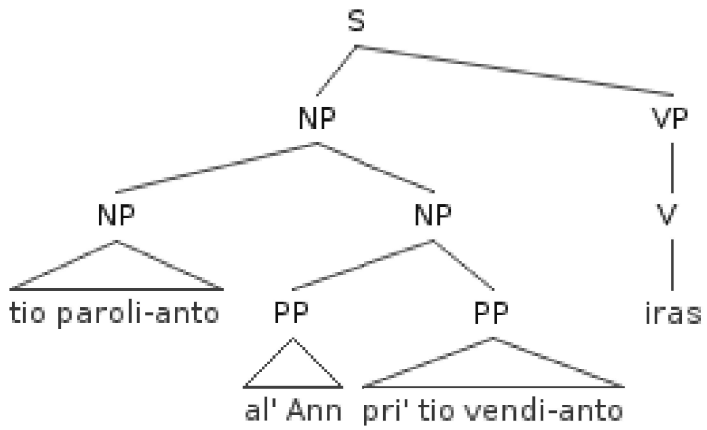

Diagram 5: as RNG exploits a natural language strategy of separating argument phrases with prepositions, no further separator particles are needed for disambiguation.

tio	paroli-anto	al'	Ann	pri'	tio	vendi-anto	iras
that	talk-subjec	to	Ann	about	that	sell-subject	go-PRED

'The talker to Ann about the seller goes.'

We have now demonstrated that an optimised model with five parts of speech (noun, adjective, verb, adverb and preposition) produces a simpler logical language than Lojban which has only predicates and particles.

To put the comparison of complexity into figures, the formal Lojban grammar consists of 1200 command lines while RNG has ten syntactic rules. Lojban has 120 different types of particles which currently add up to 1,091 individual entries. Particles make up almost half of the basic vocabulary (3.2). In contrast, RNG has no need for particles because all function words can be derived from content words.

What is more, Lojban's list of just 1,342 basic predicate words includes 3,388 separate rules for argument places while RNG employs just six rules to create a universal dictionary. Hypothetically speaking a dictionary with an infinite number of root words still requires only the same six rules for RNG; for Lojban, 2.366 * n valency rules are to be anticipated.

	syntactic rules	particle classes	total number of particles	vocabulary rules per 1,000 words	maximum number of vocabulary rules allowed
RNG	11	0	0	6	6
Lojban	1200 command lines	120	1091	2366	∞

Table 2: a comparison of the grammatical complexity of two syntactically unambiguous systems: RNG and Lojban.

7. CONCLUSION

We have examined the subject in detail and found an answer to each of our research questions (section 5):

(i) By constructing RNG, an ad hoc model based on natural language grammar, we have demonstrated that syntactically unambiguous language does not have to be based on predicate logic.

(ii) By comparing RNG and Lojban, we have demonstrated that a highly less complicated unambiguous grammar can be made using parts of speech (noun, adjective, verb, adverb and preposition) – found in many natural languages – as opposed to predicate words and particles. This finding suggests that a reductionist approach provides a significantly more optimal basis for a speakable logical language than predicate logic. Further research is however needed because neither the RNG model presented in this paper nor Lojban is optimised into a minimalistic model.

(iii) The complication of having word-specific valency rules can be avoided using prepositional phrases instead of set argument places. In addition, the complications caused by Lojban's disambiguation (bracketing) particles are avoidable.

Two previous explanations have been discussed concerning the main question which we have reformulated in the following way: What differences between Lojban and natural languages make the latter but not the former usable for the average person?

We have brought forward evidence concerning Goertzel's suggestion that the learnability problem could be solved by expanding the body of existing vocabulary. On the basis of our analysis of the complexity of Lojban's grammar, we have demonstrated that adding words makes the language more difficult to learn because much of the grammar is in the default argument places which will need to be specified word by word.

On the other hand, replacing some of the vocabulary with English root words, as proposed by Goertzel, could make language acquisition easier. It could on the other hand be a step away from Lojban's effort to have a grammar based on traditional predicate logic.

The core of our research was in providing evidence against Okrent's suggestion that the requirement of having to formulate unambiguous sentences is a fundamental obstacle for acquiring fluency. To examine this argument we constructed an ad hoc model based on natural language grammar. In comparison with Lojban, our model (RNG) proved to be hundreds of times simpler.

Our conclusion is that although natural languages tolerate a degree of syntactic ambiguity, having an unambiguous grammar per se is not the answer to the question why Lojban is not usable for the average person. Our claim is based on the evidence that a very simple unambiguous grammar can be extracted from the grammar of natural language.

This evidence could point to a design flaw in Lojban, but it seems unlikely because, as pointed out by Cowan (1995), Lojban is the result of decades of work – computer-assisted to the point of having its grammar as a computer file – by hundreds of participants, based on modern theories of language and communication.

This leads to the question whether there could be something in the underlying theoretical assumptions causing Lojban to appear complicated and unnatural. The answer is yes – provided that the creators of Lojban assumed that predicate logic would be an optimal foundation for a speakable logical language.

Our findings suggest that such a foundation would instead have to operate on a different logical mechanism, using parts of speech and constituents which can be extracted from the grammar of natural language with just a handful of rules.

References:

Arnold, Matt, 2006. *Intensive Course of Spoken Lojban.* <http://lojban.livejournal.com/18382.html>

Brown, James Cooke, 2008. *Loglan 1: A logical language*. Gainesville: The Loglan Instititute. Pdf edition <http://www.loglan.org/Download/Loglan1.pdf>

Cowan, John Woldemar, 1997. *The Complete Lojban Language. Fairfax, Virginia: The Logical Language Group.*

Eppcott, 2008. How many fluent speakers of Lojban are there? Forum post. <http://www.lojban.org/tiki/tiki-view_forum_thread.php?topics_offset=1&forumId=5&comments_parentId=7187>

Goertzel, Ben, 2006. Lojban++: An Efficient, Minimally Ambiguous, User-Friendly Natural-Like Language for Human-Computer, Computer-Computer and Human-Human Communication. <http://www.goertzel.org/papers/lojbanplusplus.pdf>

Goertzel, Ben, 2013. Lojban++: An Interlingua for Communication Between Humans and AGIs. *Artificial General Intelligence* (Lecture Notes in Computer Science) 7999. 21-30.

Goertzel, Ben, Cassio Pennachin & Nil Geisweiller, 2014. Communication Between Artificial Minds. In (ibid.): *Engineering General Intelligence*, Part 2 (Atlantis Thinking Machines 6) 411-421. Paris: Atlantis Press.

Holeš, Jan, 2004. A Priori Artificial Languages (review). *Language* 80(2). 351-352.

Kiselman, Christer, 2008. Esperanto: Its origins and early history. In Pelczar, Andrzej (ed.), *Prace Komisji Spraw* Europejskich PAU. Kraków: Polska Akademia Umiejętności. 39-56.

Libert, Alan, 2000. A priori artificial languages. München: Lincom Europa.

Logical Language Group, the, 1992. *Diagrammed Summary of Lojban Grammar Forms with Example Sentences*. <http://www.lojban.org/publications/brochures/lojex.txt>

Logical Language Group, the, 1994. *Gismu list*. <http://www.lojban.org/publications/wordlists/gismu.txt>

Logical Language Group, the, 2002. *Cmavo list*. <http://www.lojban.org/publications/wordlists/cmavo.txt>

Logical Language Group, the, 2014. *Lojban Introductory Brochure. Lojban – The Logical Language*. <http://www.lojban.org/tiki/Lojban+Introductory+Brochure>

Montague, Richard, 1970. Universal grammar. In Montague, Richard, 1974 Formal Philosophy. New Haven: Yale University Press. 222-246.

Okrent, Arika, 2009. *In the land of invented languages.* New York: Spiegel & Grau.

Okrent, Arika, 2013. Interview on Slate Radio. Bob Garfield & Mike Vuolo (ed.): *Lexicon Valley No. 33: The End of Ambiguity*. <http://www.slate.com/articles/podcasts/lexicon_valley/2013/12/lexicon_valley_lojban_is_an_invented_language_that_aims_to_be_supremely.html>

Parsons, Terence, 1990. *Events in the semantics of English*. Cambridge: MIT Press.

Powell, Robin Lee, 2010. *Lojban: You're Doing It Wrong.* Essay. <http://teddyb.org/robin/tiki-index.php?page=Lojban%3A+You%27re+Doing+It+Wrong>

Sampson, Geoffrey, 1999. The Complete Lojban Language by John Woldemar Cowan. *Journal of Linguistics* 35(1). 447-448.

Smith, Perry, 1991. Loglan 1: a logical language. Fourth edition (review). *The Journal of Symbolic Logic* 56:4. 1494-1495.

Wennergren, Bertil, 2007. Pri la *na*-ismo. *Lingva kritiko.* <http://lingvakritiko.com/2007/08/21/pri-la-na-ismo>