Journal of Universal Language 3
March 2002, 133-167

Folk Functionalism in Artificial Languages: The Long Distance Reflexive vo å in Lojban

Nick Nicholas
University of Melbourne, Australia

Abstract

A notion which underlies much functionalist thinking on language is that language is a system whose structure is engineered to solve problems in communication. Artificial languages are of particular interestinthisregard,because such problem solving can be undertaken consciously on the part of both language planners and (to the extent that the language community allows it) language users, enabling the linguistic structure to adapt to their communicative needs. Such languageusersare applying lay intuitions about whatlinguistic features will b e more effective in communication what might be characterised as ‘folk functionalism’.

An instance of such adaptation is considered here: the Lojban pronoun *vo’a*, intended as a generic reflexive, has become a long distance reflexive in order to align with Lojban’s idiosyncratic pronominal system. In fact, this seems to have beendone independently by the language planner and the language community. That the solution

yielded is typologically unusual demonstrates that communicative and paradigmatic pressures can trump natural language habit, and even typological universals in a 'perturbed' grammatical system.

1. Folk Functionalism

There are two dominant paradigms of looking at language. The *formalist* paradigm, which includes much of mainstream contemporary syntax, holds that language is to be investigated as a formal system in and of itself, and that explanations for why language is the way it is should be sought internally to that formal system. The *functionalist* paradigm, ontheother hand, seeks to explain features of language with reference to factors outside the system-namely, the function to which language is put, communication. According to such thinking, features of language are as they are because they serve to optimise communication;for instance, genderwould be explained notas a genetically coded parameter of the grammar, or an incidental feature of the lexicon, but as a mechanism for allowing the listener to keep track of the referents in a discourse.[1]

Functionalism thus implicitly regards language as a kind of problem solving, though of course opinion varies as to how this problem solving is brought about. Some functionalists favour an evolutionary approach to the function form interplay, whereby mechanisms of evolutionary selection select between forms of varying communicative efficacy (e.g. Croft 2000). Other linguists, like Scott (1985) and Hagège (1993), emphasise the deliberative contribution of individual speakers in manipulating language forms, rather than explaining language change only as an impersonal, 'invisible hand' process.

The extent to which language change characterised as 'natural' is

[1] For the latest in the long running debate between formalists and functionalists, see Darnell et al. (1998).

guided by deliberate choices is hard to gauge. But of course, language is routinely subject to change through deliberate choice in the form of prescription, which can at times have rather drastic effects on language (see e.g. the instances listed in Jahr 1989).Prescription ismotivated by various extralinguistic factors-snobbery not the least among them. Yet often the rationales invoked for a prescriptive choice appeal to notions of 'logic', clarity, and disambiguation. For example, throughout the Balkans the literary registers of languages prefer the declinable to the indeclinable relative pronoun, particularly in marked roles such as indirect object (Albanian: Buchholz & Fiedler 1987:301; Serbo-Croatian: Gallis 1956:178, Golab & Friedman 1972:43; Modern Greek: Householder, Kazazis & Koutsoudas 1964:92-93); the rationale offered for this preference is 'clarity'. So for example Papazafiri (1994:67) in a popularised prescriptive manual on Modern Greek:

> When [indeclinable] *pou* corresponds to a prepositional phrase, more attention needs to be paid to the expression. It often creates such lack of clarity, that it should be substituted by the [declinable] pronoun *o opios* and the appropriate preposition.

But the deliberate choice by a prescriptivist of a given linguistic form over another, in order to facilitate communication, is not different in essence to what functionalists claim language speakers do,when they favour a given form for functional reasons. Admittedly, the context in which prescriptivists operate is anomalous, relative to natural language change: prescriptivists are primarily concerned with written language, which as a communicative system has much less redundancy than spoken language; sothefunctionalpressures they take into account can be quite different to those of unmarked language use.

Nevertheless, if there is any validity to the notion that at least some 'natural' language change is deliberative, then prescription may yield some insights into how such change is decided. Though prescriptivists are literate and familiar with traditional grammar, they are usually

naive as far as modern linguistic theory is concerned; so the insights they have on communicative efficacy, and on where ambiguity may eventuate, may resemble what ordinary language speakers have in mind when they manipulate language. Even if it turns out that no such deliberative language change occurs 'naturally', the naive (i.e. prescientific) notions people have about language and ambiguity can help us formulate a more psychologically realistic model of how people cognitively deal with language, when they reason about it in the abstract. By analogy with 'folk psychology' as a description of people's prescientific model of the mind, I describe this kind of thinking about language, and its adaptation to communicative pressures, as 'folk functionalism'.

The prescription of literary languages is part of a spectrum of language planning; and the extreme point of that spectrum is represented by artificial languages. Since in most cases the creators of artificial languages are enthusiastic amateurs rather than professional linguists, the decisions they make as to which linguistic features to incorporate into their languages are also instances of folk functionalism. The classic instance of this, as far as the International Auxiliary Language (IAL) movement is concerned, are the recurring polemics as towhether the accusative of Esperanto is a Good or a Bad Thing.From a purely linguistic viewpoint, the question is moot: case is one way of tracking who is doing what in a discourse, word order is another, and context a third; languages successfully employ any one of these mechanisms, and it is meaningless to ask whether case or word order is more efficacious. Moreover, other factors cloud the discussion in interlinguistics: aesthetics, learnability, linguistic diffusion, etc. None the less, the question of whether a linguistic feature like case is efficient and effective in communication is certainly debated in terms of folk notions of clarity and ambiguity--i.e. in terms of folk functionalism.

Instances of folk functionalism abound in artificial language design, particularly in artificial languages intended as IALs.Attimes,theycan

be very much prescientific; the dream thanks to which Ludovik Zamenhof decided to include definite articles in his language (Boulton 1960:14) is one of the more spectacular demonstrations of this. But the creative decisions of language creators are tied up with personal preference, if not whim, and by definition cannot be tried out in usage before they are introduced. A rather more interesting phenomenon occurs when the language is released into a user community, and a communicatively inefficient or ineffectual feature is repaired by that community, applying principles of folk functionalism.

The way such change takes place in artificial languages is idiosyncratic. Artificial languages tend to be strongly prescribed, in order to forestall the language splitting up into variants (*the disdialektigho* frequently warned against in Esperanto). As Manders (1950:61) points out, the 'democratic norm' of linguistic correctness is inapplicable to Esperanto:

> In contrast with ethnic languages, in which generally only what is generally used is correct, in Esperanto one can use any expression which is comprehensible and does not contradict the *Fundamento* [language definition]. Even if all Esperantists said *Anglio* 'England' or *stulta* 'stupid' [newerforms], I would speak correctly in using *Anglujo* and *malsagha* [the original forms].

This conservative prescriptivism tends to be policed vigorously in artificial language communities; so the extent to which such languages can be altered inuseatalliscircumscribed. Indeed, in literary artificial languages (Tolkien's languages being the best instance), the language is so strongly bound to a defining canon-by community consent (not to mention legal constraint)-that it is meaningless to speak of language change carried out by the community: the community simply will not allow it.

Furthermore, for such repair to take place, the language needs to have a sufficiently large and autonomous community, to enable a

response to the communicative use of the language. A community of one (as has been the case with any number of artificial languages) is not sufficiently large to count as such a community. Likewise, change driven primarily by ideological rather than communicative pressures does not necessarily shed light on whether the alternatives are considered to be communicatively, and thus linguistically, more effective.Much of the vigorous debate on reform projects forVolapük, Esperanto, and Ido, for example, can be dismissed as ideologically motivated.

That said, there have been indisputable instances of therapy practiced on artificial languages by language communities. The strong reliance of artificial languages on written communication, and the relatively small size of their communities, mean such language change-as-therapy is more akin to the folk functionalism of natural language prescriptivists, than the obscure forces driving the 'normal' evolution of natural languages. Proponents of such changes, none the less, are able to articulate concerns about ambiguity or inefficiency in the language, and to elaborate solutions to those problems which do not fall afoul of the languages' prescriptive canonthereby preserving continuity in the languages, and drawing approval from the normally conservative community. For Esperanto, the most prominent instance has been the long drawn out search for a distinct agentive preposition, ending up with Grosjean-Maupin's *fare de/far* replacing *de* (Kalocsay & Waringhien 1980:203). For Klingon, one might mention the use of the topicaliser to disambiguate the relative clause head (Krankor 1992), which was ultimately sanctioned by the language creator (Shoulson 1995).AndLojbanhashadseveralinstances, one ofwhich is discussed here.

These proposals can be traced to particular individuals,whowielded considerable authority in the language community. (Emile Grosjean-Maupin was a lexicographer and had the ear of the editors in SAT, a sizeable organisation of Esperantists; 'Captain Krankor' is the Klingon Language Institute Grammarian.) This means that the

difference between these proposals and outright reform proposals is only one of degree; the alternative formulations are proposed because an individual senses there is something wrong with the existing language system. In order to shed light on what language users in general regard as tolerable and intolerable ambiguity in a language system, such proposals are interesting only inasmuch as they are taken up by the majority ofspeakers (which both these instances have been), and where the putative problem in the language is articulated cogently by the proponents of the formulations.

Such therapeutic changes are particularly interesting if they contravene an established tendency of natural languages, particularly the natural languages forming the substratum of the given artificial language community. The tendency to calque expressions from the substratum is demonstrably strong in second language communities (language contact providing repeated instances of this); so any assertion of linguistic autonomy on the partoftheartificial language community is of interest. Where the change runs contrary to an overall tendency of human language (as I believe applies in the case considered here), it shows that speakers' understanding of language as a system-even pathological systems like artificial languages-outweighs their tendency to follow established patterns for particular subsystems of language: if either the language community or the specific paradigm involved is anomalous, the community's use of the feature will be adjusted accordingly. Both these anomalies obtain in the Lojban pronominal system, with radical consequences for the Lojban reflexive pronoun *vo'a*.

2. Lojban

Lojban (Cowan 1997) is an artificial language based in the first instance on predicate logic. It is derived from Loglan[2] (Brown 1989),

[2] Often called by Lojbanists 'Institute Loglan' or 'TLI Loglan', after *The Loglan*

a language designed by James Cooke Brown and intended to serve as a testbed for the Sapir-Whorf hypothesis (so described in Brown 1960). The motivation of those involvedwith Lojban varies. Some are motivated bywhat they regard as the mind expanding traits of a language radically different from natural languages in its worldview. Others are interested informal logic orformalsemantics, and wish to use Lojban as a formal model of language. Yet others are interested in Lojban-which has a machine parsable syntax and a semantics formalisable to at least some extent-as a medium for human-computer interaction.

The language has been promulgated by a group of enthusiasts distinct from Loglan since the mid '80s, as the result of a dispute over the extent to which Loglan was in the public domain. That group has been led since the outset by Bob LeChevalier, who was involved in authoring much of the initial design of the language. The language had crystallised into its modern form by 1991, although refinements continued to be made to the grammar and lexicon through the early '90s, as the language started to be used more widely. The Logical Language Group (LLG), the organisation charged with developing and promoting the language (and led by LeChevalier), is committed to allowing the language to evolve 'naturally'; but it also wishes to enforce stability at least at the initial stages of the language, by imposing a baseline on the language. According to this, no proposals for reform to the language will be entertained for the first five years after the complete publication of the language. The publication of the reference grammar (Cowan 1997) has largely put a brake on revisions to the language, and the basic grammar and the lexicon have been stable since. While discussion of improvements to the language (a perennial feature of artificial language communities) continues, this deals more with features of the language underspecified in the existing language definition materials, than with revising established decisions.[3]

Institute, which administers it; Lojbanists claim their language to be a version of Loglan.

The language community primarily interacts online. Its major vehicle of interaction has been an electronic mailing list since 1989[4]; this has been supplemented more recently by Internet Relay Chat, and in 2001 the creation of a collaboratively authored Web resource, the Lojban Wiki.[5] There are also individual web pages of Lojbanists, and occasional brief face-to-face discussions in the language, particularly at the LLG's annual meeting. It is difficult to gauge the size of the language community; the mailing list as of this writing has around 250 subscribers, of which some 30 are active, and ten to fifteen regularly post in the language.

One of the primary 'selling points' for the language is that it eliminates or reduces certain kinds of ambiguity (primarily syntactic) normally inherent in natural language:

> Lojban has an unambiguous grammar (proven by computer analysis of a formal grammar with YACC), pronunciation, and morphology (word forms).... You can be very specific, or you can be intentionally vague. Your hearer may not understand what you meant, but will always understand what you said. (http://www.lojban.org/files/brochures/lojbroch.html)

The language community contains a disproportionate number of computer professionals and academics.Thelanguage community is thus more self conscious about issues of language logic and ambiguity than is typical even among artificial language enthusiasts (see Manders' (1950:61-62) comments on the applicability of the 'logical norm' to Esperanto.) As a result, there is a widespread expectation that Lojban

---

[3] For example, there has been extensive discussion through 2001 on the proper interpretation of the implicit arguments of quality nominalisations.

[4] A Listserv mailing list from 1989 to 1998, archived at http://nuzban.wiw.org/archive/; a OneGroups/YahooGroups group since 1998, at http://groups.yahoo.com/group/lojban/.

[5] Http://nuzban.wiw.org/wiki.

maintain a level of unambiguousness much in excess of what is typical for natural or artificial languages. This has consequences for the meanings assigned to the language's anaphor paradigm.

3. The Prescriptions

The Lojban anaphoric system was designed as a reaction to the Loglan anaphor system, which ithas been argued (Zwicky 1969) is not natural-that is, not conforming to the expectations that speakers of natural languages might have for an anaphor paradigm. As a compromise, however, Lojban's own anaphor system has met with disapproval among its users, both because it preserves some of the unconventionalities of Loglan, and because, paradoxically enough, users feel it compromises too much with natural language expectations. This makes of the paradigm a 'perturbed' system,which leads to unexpected results when a 'reflexive' anaphor interacts with that paradigm.

3.1. The Loglan Pronominal System

Loglan, and Lojban after it, have a plethora of pronominal forms used to refer to various kinds of referents. The subsystem of particular interest here, for its interaction with reflexives, is that of anaphora: the pronominal forms used to index a referent in the same discourse, and in particular the same clause.

In the absence of grammatical gender or number to allow the referent tracking of anaphora, Loglan has two sets of anaphors that can be employed. The first are *da, de, di, do, du*, which at any point in the discourse can be used to refer to the immediately preceding, second last, third last etc. nominal in the discourse, and retain that reference for the remainder of the discourse-not being available for reassignment to another referent (Brown 1989:173-176):

(1) *La Djank, pa vedma taj le fumna$_i$. Dai pa donsu la Pit, de$_j$. Dok pa mercea da$_i$.*
John$_k$ sold that$_j$ to the woman$_i$. X$_i$ gave Pete Y$_j$. W$_k$ married X$_i$.[6]

The second strategy Loglan employs involves acronyms: the name of a letter is used to index a nominal starting with that letter (Brown 1989:178-182):

(2) *La Tam, merji le kicmu. = Tai merji le kicmu.*
Tom is married to the doctor. = 'Tee' is married to the doctor.

In some ways, the Loglan system is underspecified; for example, where a nominal is embedded within another nominal, do they count as one or two nominals-and if two, which comes first? Moreover, counting referents whenever one wishes to anaphorise an expression becomes quickly cumbersome. And although some natural languages do have something approximating referent selection by proximity (switch reference, or the use of proximal and distal demonstratives as anaphora), a system involving such explicit referent counting is decidedly unnatural. In fact, one can argue it violates language universals, since it is context sensitive.

3.2. The Lojban Pronominal System

As an outcome of controversy over the language design being in the public domain, Lojban design has tended to be more rather than less fully specified. Moreover, in line with the premise that Lojban should not impose metaphysical constraints on thought (as a testbed for the

[6] At the end of the second sentence, *di* is the first available anaphor, and refers back to the first available nominal not already anaphorised, *la Pit* Pete; thus, *do* refers to the second such nominal, *la Djan* 'John'.

Sapir-Whorf hypothesis), it has tended to be inclusive of linguistic features, rather than exclusive.

The acronym system of Loglan is retained by Lojban (Cowan 1997:420). The other Loglan anaphor system uses a set of anaphors with permanent reference assigned by position; Lojban splits this set into two new sets, one with permanent reference, and one assigned by position, but with temporary reference.

The permanent reference anaphora, *ko'a, ko'e, ko'i, ko'o, ko'u*, are assigned explicitly to their referents by being linked with the particle *goi*, which might be glossed as 'hereafter' (Cowan 1997:150-151). Both acronyms and permanent anaphora may be observed in the following:

(3) *.uu .ue .i lo gugrxarxentinai goi koai .i baapei lo koai* ***t****urnij ba xruti le seldejni .i do stidi ma* ***t****yj.*
(Pity!Wonder!) Argentinai, hereafter X_i. Do you expect that X_i's rulers$_j$ will return the debt? Whatwould you suggest to r_j? (http://groups.yahoo.com/group/jbosnu/message/388; 'xod', 2001-12-22)

This strategy obviates counting referents in order to assign a permanent anaphor; but in turn it forces the assignment to be foreplanned: the language user needs to identify a referent likely to be talked about frequently, and assign a *ko'V* anaphor to it. Even if done as an afterthought (*.i ba'ape'o lo lo gugrxarxentina* ***goi ko'a*** *turni*), the assignment itself is highly marked.

The positional, temporary anaphors in Lojban are the set *ri, ra, ru*. Of these, *ri* refers to the immediately preceding nominal, counting by the start of the nominal phrase. (An embedded nominal is considered as following the nominal phrase it is embedded in.) On the other hand, *ra* and *ru* index 'recently used' and 'used long ago' nominals. Unlike *ri*, their scope is left deliberately vague; *ra* can in fact refer to the immediately preceding nominal, unless *ri* has already been used. (Cowan 1997:152-154). Thus, the Loglan text in (1)would be rendered as follows in Lojban:

(4) *la djan*$_k$*. pu vecnu ta*$_j$ *le ninmu*$_i$ *.i ri*$_i$ *dunda ra*$_j$ *la pit. .i ru*$_k$ *spebi'o ra*$_i$*.*
John$_k$ sold that$_j$ to the woman$_i$. X$_i$ gave Pete Y$_j$. W$_k$ married X$_i$.[7]

The Lojbansystem is certainly richer than that of Loglan; its 'vague referent' anaphors, *ra* and *ru*, are much closer to normal natural language anaphors, while the scope of the proximal anaphor *ri* is well defined, and the permanent anaphors, intended for referents persistent in discourse, are no longer contingent on the order they have been expressed in.

Nevertheless, there has been widespread dissatisfaction with the Lojban anaphor paradigm. In fact, though initially deprecated,[8] acronyms have been taken up again in Lojban, in order to compensate for the difficulty in using both the *ko'V* and the *rV* series.[9] The

[7] The literal equivalentofthe Loglan *dopamerceada*, namely *.i ru spebiori*,would have Johnmarrying himself, since *ri* refers to the nominalimmediatelypreceding it without exceptionin this instance, *ru=la djan* John.

[8] In the initial 22 draft lessons for the language, written around 1990 (http://www.lojban.org/files/roadmap.html#draft-textbook), *ko'a* is mentioned atLesson1.3, and *ri* at Lesson 21.5; acronyms are promised, but never actually mentioned. LeChevalier noted in 1993 (http://nuzban.wiw.org/archive/9312/msg00227.html; 19931017) that:

> While Lojban supports it, few Lojbanists make use of *lerfu* (letter name)words as *sumti* anaphora (back-referencing pronouns); it is coming to be heavily used in TLI Loglan, since they have realized the weakness of their other form of anaphora, which is the equivalent of our *ri*/*ra* (but with 5 members to the set, all strictly counted).

[9] Now we have re-realized, as JCB [James Cooke Brown] did, that lerfu [letters] as anaphora may be more suitable: http://nuzban.wiw.org/archive/9412/msg00321.html; Bob LeChevalier,1994122.See also,ontheLojbanWiki,Mark Shoulsons polemic

difficulty with *ko'V*-the need to foreplan their use-has already been mentioned. The difficulty with *ri* is that the strict backwards counting of referents can yield counterintuitive results. In particular, it refers to a preceding embedded nominal, rather than the preceding embedding nominal-which is more salient, being an argument of a clause:

(5) *.i mi viska {le {la djim}$_i$ patfu}$_j$ .i ri$_i$ gleki.*
I see Jimis father$_j$. He$_i$ is happy.

The difficulty with *ra* and *ru*, on the other hand, is properly not linguistic at all (since *ra* is the closest Lojban has to natural language anaphors like *he*), but ideological: Lojban has a lot invested in being an unambiguous language, so Lojbanists prefer their anaphora to be unambiguous in reference (see quotations at the end of the paper). This attitude is an aberration in human language, given how tolerant of ambiguity natural language anaphora are; and that aberration is a result of the peculiarmakeup of the community, and their expectations of the language. As John Cowan has noted to me privately,

> They don't like it [*ra*] because they are computerniks... who are obsessed with overprecision. Or in And [Rosta]'s case, a formal linguist, who is a computernik manque.

The following table shows the use of anaphors in Lojban across time:[10]

---

page, http://nuzban.wiw.org/wiki/index.php?lerfu%20pro-sumti%2C%20and%20why%20ko%27a%20sucksa very cogent articulation of folk functionalist thinking.

[10] *LL 92-93*: Lojban mailing list, 1992-6-30 to 1993-09-03. *LL95*: Lojban mailing list, 1995. *jbosnu*: Lojban languageonly mailing list (http://groups.yahoo.com/group/jbosnu,20002001). *Alice*: ongoing collaborative translation of *Alice in Wonderland* (http://www.digitalkingdom.org/cvsweb/lojban/translations/alice/). *IRC*: InternetRelay Chat logs, #lojban (http://www.miranda.org/~jkominek/lojban/; 2001-03-28 to 2001-11-28).

Table 1. Anaphor Usage in Lojban

	LL 92-93	LL 95	jbosnu	Alice	IRC
ri	49	173	23	17	45
ra	25	58	7	16	26
ru	2	1	0	0	0
ko'a	164	453	75	28	72
ko'e	34	132	6	5	9
acronyms	1	261	82	900	139

In Givón̆ian terms (e.g. Givó 1983), *ko'a* anaphors are intended for referents marked for high persistence, and *ri/ra* for referentsmarked for high locality. It seems that this split in anaphors, alien to the substratum languages, has been largely rejected by the language community, and acronyms, as anaphors allowing both types of referent, have been adopted as the preferred strategy. Indeed, prominent Lojbanists are on the record as rejecting both, in favour of acronyms:

> I think all counting anaphora in Lojban are unusable. I dont think I am able to work out on the fly (in the spoken language) which *selbri* [*predicate*] or *sumti* [*argument*] is supposed to be referred to by any of the counting or place anaphora. I do reflexives with *sevzi* ['self'] and only use *lerfu* [*acronym*] pronouns (Jorge Llambí as, http://nuzban.wiw.org/wiki/index.php?Why%20the%20Book%20is%20Right%20and%20the%20ma%27oste%20is%20Wrong)

> Just like we can use any string of *lerfu* as a *ko'a*-style *sumti* variable (which makes me think that there's practically no reason ever to use the *ko'a* series at all) (Mark Shoulson, http://nuzban.wiw.org/wiki /index.php?Type%204%20fu%27ivla).

However, more important for our purposes is the continuing use of *ri*. Lojban has available in *ri* a proximal anaphor, which as Lojban anaphors go is phonologically unmarked, and which is quite free to refer back to immediately preceding nonanaphoric nominals-with the only proviso that the noun phrase be completed.[11] Thus, the following phrases are acceptable in Lojban:

(6) a. *.i le datni$_i$ ze'e stali .i ji'a ri$_i$ ka'enai canci.*
The data$_i$ stays remains there constantly. Also, it$_i$ cannot disappear. (http://groups.yahoo.com/group/jbosnu/message/6; 'xod', 2000-1-25)

b. *i salci fa ro na'ebo le mlatu$_i$ a le ri$_i$ speni.*
Everyone celebrates, other than the cat$_i$ and its$_i$ mate. (http://groups.yahoo.com/group/jbosnu/message/149; Jorge Llambías 2000-10-10)

c. *i xu do xusra le du'u le tolkrici$_i$ cu se dimna le nu ri$_i$ krici le duu ri$_i$ na zasti.*
Are you saying the disbeliever$_i$ is doomed to Ø$_i$ believe she$_i$ does not exist? (http://groups.yahoo.com/group/jbosnu/message/154; Jorge Llambías 2000-10-29)

d. *le toknu$_i$ ri$_i$ lumci.*
The oven$_i$ cleans itself$_i$. (http://nuzban.wiw.org/archive/9505/msg00067.html; Dylan Thurston 1995-5-15)

e. *i la daos$_i$. cu seltru ri$_i$.*
The Tao$_i$ is ruled by itself$_i$. (http://nuzban.wiw.org/archive/9512/msg00169.html; Jorge Llambías 1995-12)

Examples (6d) and (6e) in particular show that Lojban already has an anaphor which (if no other nominals intervene between it and the

[11] Since a relative clause is considered part of its head nominal, for example, *ri* within a relative clause cannot refer to the head of that clause (Turner & Nicholas 2001 Chapter 9).

'subject') can serve as a reflexive. The question then becomes, what happens when an explicit reflexive is introduced into the Lojban paradigm.

3.3. The Lojban Reflexive

Lojban has two strategies dedicated explicitly to expressing reflexives. The first involves the predicate *sevzi* 'self' (in affix form, *-sez-* or *-se'i-*):

(7) a. *lu «ko'a pritu je zunle» li'u se smuni ledu'u ge ko'apritu gi ko'a zunle kei noi sevzi natfe.*
'Left-and-right' means that both X is to the right and X is to the left, which is a contradiction ("negates self"). (http://nuzban.wiw.org/archive/9208/msg00096.html; Iain Alexander 1992-08)

b. *i ny catlu la melnak noi za'o sezlu'i ni'a le farlu djacu ku'o.*
N looks at Melnak, who washes himself ("self-washes") too long under the shower. (http://groups.yahoo.com/group/lojban/message/3495; Jorge Llambías 2000-7-9)

This strategy is properly lexical rather than anaphoric, and is not restricted to reflexivisation.[12] Creating a new lexical compound to bring about reflexivisation is felt to be overkill for a variety of reasons (see discussion in http://nuzban.wiw.org/archive/9505/msg00067.html); so this strategy is deprecated, and is not properly in competition with ri as a reflexive.

On the other hand, the anaphors in the series *vo'a, vo'e, vo'i, vo'o, vo'u* (Cowan 1997:158-159), indexing the first, second, third, and so on place of the predicate, are ostensibly meant as reflexives-specificaly

[12] For example, Jorge Llambías uses *le sevzi mukti* in http://nuzban.wiw.org/archive/9504/msg00044.html to mean own motives or selfish motives.

vo'a, which indexes the first place of the predicate, and is thus equivalent to the normal natural language reflexive, indexing the subject. Thus, the following expressions are possible in Lojban:

(8) a. *mi*$_i$ *lumci vo'a*$_i$.
I$_i$ wash myself$_i$.
b. *mi*$_i$ *dunda le cukta vo'a*$_i$.
I$_i$ give myself$_i$ the book.
c. *.ifinti fa la cevni*$_i$ *loi remna goi fo'aneta'iletarmibevo'a*$_i$.
God$_i$ created Man (hereafter Y) in His$_i$ own image. (http://nuzban.wiw.org/archive/9201/msg00052.html; Mark Shoulson 1992-1-22)

Vo'a is also used in the reciprocal construction *soi X Y*, meaning that the current predicate is still valid if *X* and *Y* are swapped. The construction usually appears as *soi vo'a*, where the second argument is ellipted if understood to be the nominal next to *soi X*. Thus, (9a) and (9b) are equivalent:

(9) a. *mi*$_i$ *prami do*$_j$ *soi vo'a*$_i$ *vo'e*$_j$.
I love you and vice versa. (= *swap* I *and* you)
b. *mi*$_i$ *prami do soi vo'a*$_i$.
I love you and vice versa; we love each other.

Unlike *ri*, *vo'a* as a reflexive can index anaphoric subjects (8a), and is not restricted to cases where it is the first nominal after the subject (8b). Moreover, it resembles natural language reflexives in that it is specifically marked for reflexivity-unlike *ri*, which is an all purpose anaphor (being marked instead for discourse proximity.) Ostensibly, this means that Lojbanists should prefer *vo'a* as a reflexive over *ri*, both because it more strongly resembles the substratum model (natural language, particularly English reflexives), and because it is more flexible than *ri* in what it may index.

The difficulties with *vo'a* beginwhenitisusedinaclauseembedded within another clause. This introduces an ambiguity thatneeds to be resolved: does *vo'a* index the first place of the embedded clause, or of the matrix clause? In natural language, the answer is clear: the majority ofreflexives in the world's languages (including English) are shortdistance, so that they index subjects in the sameclauseasthereflexiveitself(10a).Longdistance reflexives, which can index the subject of the matrix clause as well as the local clause (Cole, Hermon & Huang 2001), are typologicallyunusual; the languages they appear in include Chinese, Kannada,Dutch, and Icelandic (10b), but none of the languages which might reasonably be considered 'substrata' to Lojban-not English, nor Russian, Bulgarian, Spanish, Hebrew, Greek, and the other second languages of prominent Lojbanists.

(10) a. *John$_i$ says that Mary$_j$ loves herself$_j$/*himself$_i$.*
b. *Jón$_i$ segir að María$_j$ elskar sig$_{i/j}$.*

This ambiguity has been reflected in the uncertainty in the definition of vo'a on thepartofthelanguageplanners-more so, as itturns out, than in the language community itself. Initially, *vo'a* could be indexed to a particular clause, using Lojban's repertoire of clausal anaphora. Thus, one could speakof *vo'a pe diu*, the first argumentofthepreceding sentence's predicate, as distinct from *vo'a pe dei*, the first argument of the current sentence's predicate. By default, *vo'a* referred to the preceding sentence.

However, this indexing did nothing to resolve the ambiguity between short distance and long distance reflexive, since *vo'a* (at the time) could be anchored only to sentences, and not to clauses within sentences.[13] In fact, by indexing nominals in preceding sentences, *vo'a*

[13] Before sentenceanchoring was introduced, *vo'a* was prefixed to predicate anaphors rather than sentential anaphors. This would have allowed a local/matrix distinction (*vo'afui*: subjectofmatrixpredicate; *vo'afai*: subjectoflocalpredicate.)This facility was eliminated before the language was published. (http://nuzban.wiw.org /archive/9407/msg00035.html; Bob LeChevalier 1994:713), so the language community never became aware of it.

was not behaving as a reflexive at all, but more like a logophor.

When this logophoric use was supplanted by the nominalisation of predicate anaphors (*le go'i* 'the [first argument of the] preceding sentence's predicate'), the reciprocal use of the anaphor made the main language planner, LeChevalier, decide on long distance reference:

> One major purpose of '*vo'a*' is for explicitly dealing with 'and vice versa' which has a special metalinguistic syntax (*soivo'evo'a*). It really WAS intended to bounce you out to the main *bridi* [*predicate*], because I didn't conceive of the need to refer to other *sumti* [*arguments*] at the subordinate level. (http://nuzban.wiw.org/archive/9407/msg00035.html; Bob LeChevalier 1994-7-13)

The conflation of reflexives and reciprocals into the same anaphor, and the language planner's folk functionalist judgement that reciprocals were far likelier to have matrix referents than local referents, made him decide to give *vo'a* long distance rather than short distance reference. The definition of *vo'a* was already established in October 1988, with the first published word list for the language:

> In counting *sumti* for *vo'a* series anaphora, you are concerned only with the *sumti* (and not modal/tense operators) of the main *bridi* of an utterance, as they are formally defined. ...For *vo'a*, this counting goes regardless of whether the nature (*sic*), and not with subordinate *bridi* that are found within *sumti*. *Vo'a* series anaphora are most useful in reflexive constructions, and in subordinate clauses that reference *sumti* of the main clause.
>
> **vo'a** **VOhA** he/she/it; pro-*sumti* representing the 1st *sumti* of a *bridi*; defaults to the main *bridi* of this utterance. (http://www.lojban.org/files/history/CMAV1088.ZIP)

The current version of the wordlist (not greatly modified since 1991) retains this definition:

> **vo'a** pro-*sumti* [*pronominal*]: repeats 1st place of main *bridi* [*predicate*] of this sentence (http://www.lojban.org/files/wordlists/CMAVO)

Since work on Lojban proper only began in 1987, *vo'a* has in fact been defined as long distance since the language was published. Moreover, the flexibility of assigning *vo'a* to different clauses has not persisted in the language; of the 112 instances of *vo'a* used in Lojban text on the Lojban mailing list from 1989 to 1998, only five instances are logophoric: three early instances (Jim Carter 1990, Bob LeChevalier 1991, Ivan Derzhanski 1992), and two instances produced by a beginner, and immediately corrected (Scott Brickner 1996).

However, there is a clash in authority between the word lists and the Reference Grammar, which is intended to be the definitive prescription for the language. In the latter, Cowan (1997:158) describes *vo'a* as follows:

> The *cmavo* [*function words*] of the *vo'a*-series are pro-*sumti* anaphora, like those of the *ri*-series, but have a specific function. These *cmavo* refer to the other places of the same *bridi* [*predicate*]; the five of them represent up to five places.

By binding *vo'a* to the predicate (i.e. the clause), rather than (the matrix clause of) the sentence, Cowan defines *vo'a* as a short distance reflexive. When the issue came up for discussion in 2001, John Cowan[14] defended his position vigorously:

[14] Cowan is vice president of the Logical Language Group, and has done the bulk of the work on Lojban grammar since 1990. He can thus be considered a language planner for Lojban almost as much as can LeChevalier. Note that Cowan came to the language after LeChevalier had decided *vo'a* should be long distance.

This page is polemical, and is written by John Cowan

> IMAO[15] the *cmavo* list definitions of *vo'a* etc. as referring to the sumti of the *main bridi*, rather than the *current bridi* (as the reference grammar has it) is bogus. Reflexives are normal and useful. Long-distance anaphora are really not.
>
> In particular, applying *vo'V* with the narrowest possible scope, even within descriptions, allows the creation of **reflexive selbri**. For example, a "self-lover" would be *le prami be vo'a*. This cannot be done with the current understanding of *vo'a*. (http://nuzban.wiw.org/wiki/index.php?Why%20the%20Book%20is%20Right%20and%20the%20ma%27oste%20is%20Wrong)

With his nonfolk knowledge of linguistics, his understanding of *vo'a* as primarily a reflexive rather than a reciprocal, and the high regard in which his grammar is held, it seems strange thatthere should be any question that Cowan is right. However, the data shows otherwise.

4. The Data

The following corpora have been examined for instances of *vo'a*: the old Lojban mailing list (1989-1998); the new Lojban mailing list (1998-2001-Michael Helsem's extensive short story, *lapoi pelxu ku'o trajynobli* 'the Yellow King', is dealt with separately); the *jbosnu* list (2000-2001); the IRC logs (2001); and *Alice* (2001) (on these, see above.) Repetitions, quotations, metadiscussions, and sentences generated by computer have been eliminated.

Instances of *vo'a* are separated into arguments of nominals and of clauses. These correspond to natural language possessive and object

[15] In my arrogant opinion.

reflexives. As Cowan has stated, taken to its logical extreme short scope *vo'a* within a nominal should refer to the head of the nominal (since Lojban recognises no syntactic distinction between nominal predicates and clausal predicates.) This would allow the parallel constructions:

(11) a. *la fred$_i$. catra vo'a$_i$.*
Fred$_i$ kills himself$_i$.
b. *le catra be vo'a.*
the killer$_i$ of himself$_i$; the suicide.

—something Cowan finds highly desirable. However, apart from one posting in 1994 proposing this turn of phrase (http://nuzban.wiw.org/archive/9407/msg00031.html, Jorge Llambí as), there has been no such usage in Lojban: *vo'a* embedded within nominals always has its reference outside the nominal, as in (11c).

c. *za'a le prenrkore'ai cu katna .oiro'odai vimcu le jipno be le degji be vo'a$_i$.*
The Korean$_i$, as I saw, cut off (ouch!) the tip of his$_i$ finger. (http://www.miranda.org/~jkominek/lojban/log-2001.11.09.txt; Arnt Richard Johansen)

The referents of *vo'a* are classed as logophoric (when the referent is outside the sentence), *long distance* (when the referent is outside the immediate clause), *matrix short distance* (when the referent is in the same clause, which is also the matrix clause of the sentence-so that long distance reference is impossible), *marked short distance* (when the referent is in the same clause, which is embedded within another clause-so that long distance reference is possible), and *ambiguous* between long and short distance reference.

The results are as follows:

Table 2. Functions of *vo'a* in Lojban

	Old List	New List	pelxu	jbosnu	IRC	Alice	Total
Clausal	73	53	23	11	15	5	180
Nominal	39	21	5	8	5	5	83
Reflexive	105	70	28	18	20	20	251
Reciprocal	7	4	0	1	0	0	12
Total	112	74	28	19	20	10	263

Table 3. Scope of *vo'a* in Lojban

	Old List	New List	pelxu	jbosnu	IRC	Alice	Total
Logoph.	5	0	0	0	0	0	5
Long.	23	35	20	9	10	8	105
Matrix	68	29	6	7	10	1	121
Short Marked	6	5	0	2	0	1	14
Short Ambig.	7	4	2	1	0	0	14
Other	3	1	0	0	0	0	4
Total	112	74	28	19	20	10	263

Although *soi vo'a* is reported by Lojbanists to be a quite fixed collocation (on which more later), the reciprocal has not been widely used: it constitutes one twentieth of all usages of *vo'a*.[16] So although LeChevalier anticipated the reciprocal to be a major usage of *vo'a*, and indeed steered its semantics according to that perception, this is not borne out in usage.

The more interesting results for our purposes are the scope of *vo'a* as an anaphor.

[16] The rationofreflexivestoreciprocals in English iscloserto1:1000(usingnumber of instances in the British National Corpus for *each other, eachother,* and *one another versus himself, herself, itself, themselves, oneself:* http://sara.natcorp.ox.ac.uk/.)

Clearly the vast majority of usage of *vo'a* holds with LeChevalier's prescription, rather than Cowan's. The instances in which Lojbanists, given the option, choose a long distance rather than short distance interpretation for *vo'a* have been numerous since the outset, and have actually increased with time. Cowan's prescription, despite coming in a much fuller description of the language than the old word lists, has had no effect on usage. One could argue that this is a result of Lojbanists not changing the usage they acquired pre-1997. Yet, although several prominent Lojbanists have been active since before the publication of the reference grammar (Jorge Llambías and Michael Helsem being the most notable), a sizeable number has come to the language only afterwards. (Jay Kominek, whose defence of long distance scope is given below, is an example.) And Lojbanists cannot be said to have slavish adherence to past usage, particularly as much of it predates important language reforms undertaken in the early '90s.

Table 4. Scope of clausal *vo'a* in Lojban

	Old List	New List	pelxu	jbosnu	IRC	Alice	Total
Logoph.	5	0	0	0	0	0	5
Long.	21	32	19	7	9	4	92
Matrix	37	5	3	2	6	0	53
Short Marked	4	4	0	2	0	1	11
Short Ambig.	3	3	1	0	0	0	7
Other	3	1	0	0	0	0	4
Total	73	53	23	11	15	5	188

The proportion of long scope usage is even greater among clausal reflexives. One can argue that in Lojban, since there is no intrinsic difference between the predicate in a nominal and in a clause, and reflexive nominal reflexives like (11b) are unused, all nominal reflexive use of *vo'a* should already count as long distance. If the counts are

restricted to clausal reflexives, the following results obtain:

What these results show is that not only is long distance usage entrenched in Lojban, but Lojbanists in fact go out of their way to use *vo'a* as a long distance reflexive-it is used almost twice as often in embedded as in matrix clauses.

The instances where an explicitly short distance scope is chosen over long distance are themselves revealing. While they have not been frequent enough to establish a clear pattern, one can see functional motivations for the violation of accepted usage in some of them. In (12a), for instance, the embedded clause is itself the subject of the matrix clause; the reflexive referring to the clause containing it would be unnaturally recursive, so the long distance reading ($vo'a_j$) is rejected in favour of short distance ($vo'a_i$):

(12) a. *{lenu la graf.*$_i$ *noi vipsi cmagu'etru cu casnu vo'a*$_i$ *lejei loi na'e gunka jdini cu dunda loi pu sonci kei kei}*$_j$ *cu rinka lenu le cmagu'e dinsro cu sorcu noda.*
{The fact that vice-governor Graf$_i$ discussed with himself$_i$ whether unemployment benefits were given to veterans}$_j$ caused the state treasury to be empty. (http://nuzban.wiw.org/archive/9203/msg00034.html; Mark Shoulson 1992-3-8)

In (12b), on the other hand, the culprit is clearly the reciprocal construction. Contrary to LeChevalier's guess, there is strong pressure for reciprocals to have short distance reference: reciprocality is apparently likelier within a predicate (*I see Greeks kiss Turks and Turks, Greeks*) than across a predicate boundary (*I see Greeks kiss Turks, and Greeks see me kiss Turks*), and the referent of *vo'a* has been chosen accordingly:

b. *.i pi'o lo veltivni ku mi*$_j$ *zgana lenu lo xespre*$_i$ *lo prenrturkie cu cinba soi vo'a*$_i$.
On television I$_j$ see Greeks$_i$ and Turks kiss each other$_i$. (http://groups.yahoo.com/group/lojban/message/1190; Robin Turner 1999-8-25)

In (12c), finally, the short scope interpretation may have been chosen either because the long distance interpretation would index the first person pronoun (which would be marked, since *mi lumci mi* is perfectly acceptable in Lojban for 'I wash myself'), or as an attempt at emphasis, calquing English ('his own country'):

c. *.ije'u mi$_j$ na zanru lenu lo se zajbrnatleta$_i$ cu gubgau lei cfila be le cecmu jditai be levo'a$_i$ jecta.*
Honestly, I$_j$ don't see why an athlete$_i$ should show off the disadvantages of the social policy of his$_i$ country. (http://nuzban.wiw.org/archive/9201/msg00062.html; Nick Nicholas 1992-1-16)

5. Discussion

Compared to 'normal' languages, Lojbanists' decision to choose a long distance scope for their reflexive--indeed, to prefer it in long distance contexts--seems perverse. But Lojban is not a normal language. In particular, its anaphoric paradigm is not normal; and in the best structuralist tradition, the semantics of *vo'a* has been determined paradigmatically, contrasted with the nonreflexive members of the anaphoric paradigm. To see how this works, let us attempt a functionalist comparison.

In natural language, the opposition betweenreflexives and unmarked anaphors can be characterised as resulting from two markedness hierarchies. First, reflexive predicates are more infrequent, and thus more marked, than nonreflexives; so a reflexive referent needs to be signalled by a marked anaphor, whereas the unmarked anaphor becomes free to index any nonreflexive referent.[17] Second, unmarked

[17] CompareCroft (1991:256) on themarkedness of reflexiveand reciprocal predicates in their linguistic encoding through voice.

anaphors are free to select for more topical referents. Referents in the matrix clause are more accessible for anaphoric reference than distinct referents in embedded clauses (Ariel 1990). A short scope referent will thus be more marked than a long scope referent, so it is encoded with the more marked reflexive anaphor.

Such accounts contrast reflexives with unmarked anaphors, which are unconstrained in what referent they can index; the accessibility of referents to such anaphors is primarily a property of discourse. Lojban has no explicit anaphors unconstrained as to their referent (although zero anaphors in the language, which are abundantly used, clearly follow such a pattern of accessibility.) *Ko'V* anaphors have fixed referents assigned to them in order to be meaningful at all. Acronyms are likewise fixed to the nearest preceding nominal starting with the given letter. *Ri* is fixed by proximity; *ra* is only slightly more free, but is not widely used in any case. Topicality is thus irrelevant in Lojban anaphora (although Lojbanists can select what they think will be topical, and assign *ko'V* to it.)

So other than zero anaphors, *vo'a* is not competing with anaphors unmarked as to their choice of referent. The choice of a short scope referent is no less marked in Lojban than the choice of a long scope referent, a proximal referent, or a preassigned referent (*ko'V*). So there is no good candidate anaphor to take up the unmarked case complementary to the reflexive.

On the other hand, the short scope reflexive, which typically involves greater proximity than long scope, already has competition in *ri* and *ra*. And the unmarked referent in embedded clauses-the matrix subject, when distinct from the embedded subjec-has no distinct anaphor available to index it. Since Lojbanists want to encode the same kinds of referents they do in their native languages, it is inevitable that the reflexive be coopted to index that unmarked referent, by having a long distance scope imposed on it. The decision was made initially by the language planner LeChevalier, though with a different stated motivation (which as seen seems to have been wrong.) Yet, by actively

associating *vo'a* with long distance contexts, rather than merely using *vo'a* regardless of the level of embedding, the language community is seen to be consciously manipulating *vo'a* to that end.

Lojbanists with no special linguistic training have articulated this train of thought explicitly; the following are representative samples:

> I've personally had need to use *vo'a* as defined by the maoste [word list], and never any need to do what The Book [Cowan 1997] says. But if I did, I could use lenei, lesenei, etc.[18] And without backcounting (not a pleasant prospect, unless you do lenoaxiro (ugly)),[19] there isnt a convenient way to refer to the sumti of the main bridi. And that is where the topic/subject of the sentence is going to be, and that is going to be one of the most commonly repeated sumti, I suspect. (http://nuzban.wiw.org/wiki/index.php?Why%20the%20Book%20is%20Right%20and%20the%20ma%27oste%20is%20Wrong; Jay Kominek)

> It seems more likely to be useful. I, at least, have encountered more uses for the ma'oste's version. The topic/subject of the predication is going to be in the main *bridi*, not one of the sub ones, and subpredications are frequently going to need to refer to the topic/subject. *Lenei* or *lesenei* work

[18] Nominalisations of the current predicate anaphor: the first argument ofthepresent predicate';the secondargumentofthepresentpredicate'.Thisanaphorhas notbeen usedmuchtodate;butLojbanistshavealreadyadvocatedusing *lenei* insteadof *vo'a* in all short range contexts, even in Lojban paedagogy (http://groups.yahoo.com/group/lojban/message/10021, http://groups.yahoo.com/group/lojban/message/10047;RobSpeer2001-8-24, 2001-8-25).

[19] *No'a* is the anaphor of the matrix of the current predicate. In cases of multiple embedding, *no'a* is subscripted: *no'a*xipa 'first predicate outside', *no'axire* 'second predicate outside',uptothe(clumsy) *no'axiro* 'allthpredicate outside', i.e. 'matrix predicate of sentence'.Thenominalisations of *no'axiro* arethus themostgeneral way ofreferring to outermostmatrix nominals, regardless ofthedepth of embeddingof the current predicate.

> in the much more unlikely case of needing to refer parts of the current sub-*bridi* [*subpredication*]. (http://www.miranda.org/~jkominek/lojban/log-2001.11.09.txt; Jay Kominek)

The kind of thinking Kominek is displaying is rather more linguistically informed than the typical call to use declinable rather than indeclinable relativisers in Balkan languages, or for that matter the accusative in Esperanto. Nevertheless, I believe it is not different in essence. And whether with conscious linguistic forethought, or through the trial and error of finding they could not anaphorise the referents they wanted to, Lojbanists have applied folk functionalism to the anaphora available to them, giving a reflexive quite different to their substratum systems, but well suited to the paradigm they have found in the language.

Moreover, the responses to a proposal by Nick Nicholas to allow pragmatic flexibility in the interpretation of *vo'a* (as a response to the actual behaviour of Lojbanists shown in the corpus: http://groups.yahoo.com/group/lojban/message/10005, 200184) drew a revealing set of protests:

> I don't like this. *Vo'a* was one of the pronouns for which it is possible to absolutely tell what its referent is; there aren't many others. (http://groups.yahoo.com/group/lojban/message/10021; Rob Speer 2001-8-24)

> OTOH,[20] doing what Nick proposes, and formalizing usage patterns into documented conventions, will serve as explicit and warning testimony to the fuckups that arise by leaving things to usage to decide. (http://groups.yahoo.com/group/lojban/message/10042; And Rosta 2001-8-25)

[20] On The Other Hand.

> This wasn't left to usage intentionally, it was a mistake. The real problem is that *vo'a* was usually intended as long-distance when alone, and usually short-distance when used with *soi*. The obvious answer is to make it long-distance when there is no *soi*, and short when there is. I want to be able to know certainly what *vo'a* means. (http://groups.yahoo.com/group/lojban/message/10043; xod 2001-8-25)

Speer's and xod's responses show that Lojbanists, for reasons of linguistic ideology, cherish nonambiguity in anaphor reference, and will resist attempts to make their referentiality more 'free' and therefore more natural-language-like. Rosta's response (particularly interesting in that it comes from a professional linguist) shows that there is enduring mistrust in at least part of the language community for the 'democratic' norm of correctness-an attitude consistent with artificial languages, as Manders found, and their prescribed literary natural counterparts, though not with the stated policy of the LLG as articulated by LeChevalier.[21]

Xod's response, lastly, concerns a problem identified by more than one Lojban folk functionalist: reciprocals and reflexives have opposing requirements in their scope (see 12b). This issue has not been finalised, although it seems likely that uniformity of definition will outpoll folk functionalism in this instance, and *soi vo'a* will also become long scope-with the introduction of the explicitly short-scope anaphor *lenei* serving to counterbalance this.

6. Language Design

Lojban is not primarily intended as an IAL-notwithstanding occasional claims of its suitability as an IAL, made within the

[21] Http://nuzban.wiw.org/archive/9703/msg00012.html, theannouncementofthe Lojban Baseline, is an *ex cathedra* articulation of this policy.

community. However, just as the behaviour of *vo'a* reveals peoples notions of folk functionalism, the therapy that artificial language users apply to misconfigurations in their language systems in general reveals what makes for sound design in artificial languages. A language whose community ends up practicing therapy on it is clearly not well-designed: the initial language design should forestall the need for any such therapy. And while robust language communities (like those of Esperanto and natural languages) can cope with language change, a fragile language community can be fatally damaged by it. The fractious history of IAL projects, and the endless disputations that have plagued them, are a cautionary tale to anyone who would venture into the field with a new project.

Ostensibly, an artificial language intended to be easily learned by humans--as IALs obviously are--should conform to the norms of what is natural in human language. And if two alternative constructions turn up in human language, the obvious choice for an IAL is the most typologically diffused construction: a cross-linguistically unmarked strategy will be more readily acquired by a greater proportion of language learners, from a wider range of language backgrounds. From this point of view, it is self-evident that any IAL with reflexives should prefer a short-range over a long-range reading: short range reflexives predominate among the languages of the world. It would seem perverse for any language community to wish otherwise, and to seek out a long range reflexive, unmotivated either by first language transference or any notion of linguistic 'naturalness'.

But this is precisely what has happened with Lojban. The language community is indeed somewhat 'perverse', given its preoccupation with explicitness and unambiguity. But as shown, the choice of a long distance reflexive is very well motivated, by the nature of the paradigm within which the reflexive is used. The inconsistency between short distance reflexives and the Lojban anaphor system has led to therapy on the language undertaken by the language community.

As the example of *vo'a* shows, decisions on what constitutes a

natural or desirable feature for an artificial language cannot be taken in isolation; linguistic features cannot simply be selected survey-style for a language intended for real use within a language community. Just as a language has to be typologically plausible to be readily learnable, so too does it have to be paradigmatically consistent, and linguistically coherent. If two subsystems of the language are natural in themselves, but inconsistent with each other, the language community will reject the combination.

Much has been made of the 'linguistic intuition' Zamenhof displayed in designing Esperanto, even though he was at times quite naive linguistically:

> According to Jespersen's definition [...] interlinguistics "is concerned with the structure and basic concepts of all languages, intending to establish a norm for interlanguages". In fact, these words give the right idea about the working methods of interlinguists: in an almost mechanical way they manipulate the grammars and vocabularies of ethnic languages, and following rigorously fixed (but insufficiently motivated) principles, they exhaustively collect all elements they believe necessary for the ideal planned language, with admirable exactitude. In this work Zamenhof and his theories cannot be of use: Zamenhof was only slightly interested in minor details, and his language often shows signs of neglect and inattention. But if interlinguists would study attentively Zamenhof's reasoning and hypothetical considerations, they might stop placing so much value *a priori* on details (Manders 1950:15).

Even allowing for the Esperanto polemic, it is clear that Esperanto is successful as a coherent linguistic system. The 'genius' of Esperanto need not be sought in mystical terms; it follows from the successful integration of the grammatical features of the language into a cohesive whole. A similar cohesion should be a desideratum of artificial

language design in general, and language therapy demonstrates the need for it, and the consequences of ignoring it.

A complementary instance of therapy in Lojban has been the resumption of the formerly deprecated acronym anaphors. In this instance, the therapy can be described in terms internal to the anaphor paradigm: the split between high locality anaphors and high persistence anaphors is indeed unnatural in terms of human language, and is being increasingly rejected by Lojbanists. While there will always be a niche for *ri*, it is not impossible that acronyms will in time displace *ko'V* anaphors. This serves as a reminder that naturalness itself is still important in language design, particularly when it involves such communicative fundamentals as referent tracking. So successful language design needs to balance the potentially conflicting imperatives of typological naturalness and paradigmatic coherence.

References

Ariel, M. 1990. *Accessing Noun-phrase Antecedents.* London: Routledge.

Boulton, M. 1960. *Zamenhof: Creator of Esperanto.* London: Routledge & Kegan Paul.

Brown, J. C. 1960. Loglan. *Scientific American* 202.6, 5363.

Brown, J. C. 1989. *Loglan 1: A Logical Language.* Gainesville, FL: The Loglan Institute, 4th Ed.

Buchholz, O. & W Fiedler. 1987. *Albanische Grammatik.* Leipzig: VEB Verlag Enzyklopädie.

Cole, P., G. Hermon, & C. Huang. 2001. *Syntax and Semantics 33: Long-distance Reflexives.* San Diego: Academic Press.

Cowan, J. W. 1997. *The Complete Lojban Language.* Fairfax, VA: The Logical Language Group.

Croft, W. 1991. *Syntactic Categories and Grammatical Elations: The Cognitive Organization of Information.* Chicago: University of Chicago Press.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach.*

Longman Linguistics Library. New York: Longman.

Darnell, M. et al. 1998. *Studies in Language Companion Series 41: Functionalism and Formalism in Linguistics.* Amsterdam: J. Benjamins.

Gallis, A. 1956. *The SyntaxofRelativeClauses in Serbo-Croatian Viewed on a Historical Basis*. Oslo: Aschehoug.

Givòn, T. 1983. Topic Continuity in Discourse: An Introduction. In T. Givòn (ed.), *Topic Continuity in Discourse: A Quantitative Cross-language Study* 1-42.

Gola‰, Z. & V. A. Friedman. 1972. The Relative Clause in Slavic. In P. M. Peranteau, J. N. Levi & G. C. Phares (eds), *The Chicago Which Hunt: Papers from the Relative Clause Festival* 30-46.

Hagège, C. 1993. *Current Issues in Linguistic Theory 94: The Language Builder*. Amsterdam: J. Benjamins.

Householder, F. W., K. Kazazis, & A. Koutsoudas. 1964. *Reference Grammar of Literary Dhimotiki.* Bloomington, IN: Indiana University.

Jahr, E. H. 1989. Language Planning and Language Change. In L. E. Breivik & E. H. Jahr (eds.), *Language Change: Contributions to the Study of its Causes* 99-113.

Kalocsay, K. & G. Waringhien. 1980. *Plena Analiza Gramatiko de Esperanto.* Rotterdam: Universala Esperanto-Asocio, 4th Ed.

Captain Krankor. 1992. From the Grammarian's Desk. *HolQeD* 1.3, 4-6.

Manders, W. 1950. *Interlingvistiko kaj Esperantologio.* Purmerend, Netherlands: J. Muusses.

Papazafiri, I. 1994. *Lathi sti Khrisi tis Glossas Mas.* Athens: Smili, 9th Ed.

Scott, J. E. 1985). Vivid Language and Language Change. In A. Ahlqvist (ed.), *Papers from the 5th International Conference on Historical Linguistics, Amsterdam Studies in the Theory and History of Linguistic Science* 21, 304-315.

Shoulson, M. 1995. Interview: Okrand on **-bogh** and More. *HolQeD* 4.2, 5-6.

Turner, R. & N. Nicholas. 2001. *Lojban for Beginners.* Available at URA <http://www.opoudjis.net/lojbanbrochure/lessons>.

Zwicky, A. M. 1969. Review of Brown, J.C., "Loglan 1". *Language* 45.2, 444-457.