

Lĩhusã documentation (English)

Contents

1	Introduction	2
2	Free phoneme order	2
2.1	Example	3
3	Background	4
4	Phonology	5
4.1	Consonants	5
4.2	Vowels	5
4.3	Syllable structure	5
5	Reduplication	6
5.1	Syntactic reduplication	6
5.2	Morphological reduplication	6
5.3	Deduplication	6
6	Morphology	8
6.1	Roots	8
6.2	Affixes	8
6.2.1	Linking particle and compound predicates	8
6.2.2	Case particles and cases	8
6.2.3	Voice markers and Austronesian alignment	9
6.2.4	Semantic affixes	10
6.3	Numerals	11
7	Sentences	12
7.1	Elision	12
7.2	Serial predicates	12
7.3	Negation & Prohibitive	12
7.4	Existence	12
7.5	Tense	13
8	Writing system	14
8.1	Case markers	14
8.2	Voice markers	14
8.3	Other affixes	14
8.4	Roots	14
8.5	Reduplication	14

1 Introduction

Lĩhusã ([lĩ'husã] or any permutation of the phonemes, lit. '(language) of the turbulent sea') is a (cursed) constructed language made by Kat Mistberg / 霧山 Kat. It is intended as a demonstration that a language with **free phoneme order** is possible: the order of the phonemes within each Lĩhusã sentence can be scrambled into any order *without loss of information*.

Other notable features are:

- **Austronesian alignment:** Lĩhusã is, after all, a Formosan language.
- **Non-linear writing system:** because the order of phonemes carries no information, it can — and usually is — written with a non-linear writing system.
- **Human-usable:** despite the above features, it is very much possible for ordinary humans to learn and use this language.

Lĩhusã won first place in the Cursed Conlang Circus 4 (CCC4):

- [CCC4 submission video](#)
- [CCC4 judges' reaction](#)

2 Free phoneme order

The most prominent feature of Lĩhusã is that within each sentence, the order of the phonemes carries no information. This section explains how this can be achieved in general.

The idea is that we first create an ordered language — a language where the order of phonemes does convey information — and then transform it into an unordered language. There are many possible such “unordering transforms”, but arguably the simplest one, and the one used in Lĩhusã, is done using repeated reduplication. Let's assume that sentences in the ordered language are ordered sequences of words, each ordered word has at most one copy of each phoneme, and ordered words are not differentiated by the order of phonemes within the word (i.e. it's never that two ordered words differ only by phoneme order). Then we label the words of the sentence, from the start to the end, with numbers 0, 1, 2, 3, The transformation is then to say the k th word 2^k times — equivalently, we reduplicate the k th word k times — to get the equivalent sentence in the unordered language. In essence, reduplication replaces word order. We call k the “reduplication count”, or “dupe count” for short.

This transformation is always invertible using the equivalent “ordering transform”. Let's say we have an unordered sentence. Since it is unordered, the only information available to the listener is the counts of each of the phonemes (the **phoneme counts**). Using the phoneme counts, the ordered sentence can be reconstructed using the ordering transform as follows:

1. Determine each phoneme's phoneme count and write it in binary.
2. For each integer $k \geq 0$, look at the digits of the 2^k 's place of the binary phoneme counts of all the phonemes. These are then the phoneme counts of the word in the sentence with dupe count k .

What makes the unordered language unordered is that no matter which order the phonemes are in in the unordered sentence, the ordering transform will always map it to the same ordered sentence, as it only takes the phoneme counts as input. The inverse transformation works because each word contains at most 1 copy of each phoneme, and the k th word is said 2^k times: the k th word only affects the 2^k 's digit of the phoneme counts written in binary, and the 2^k 's digit can only be affected by the k th word. We thus have the following result:

If each word in the language contains at most 1 instance of each phoneme, and each word in a sentence is reduplicated a different number of times, then no matter in what order you place the

phonemes in the unordered sentence, it is always possible to reconstruct which words appear in the ordered sentence and how many times each word has been reduplicated.

2.1 Example

Let's say we have the ordered sentence mxad to+ixu onia . To perform the unordering transform, we reduplicate the first word 0 times (i.e. say it 1 time), the second word 1 time (i.e. say it 2 times), and the third word 2 times (i.e. say it 4 times), so we get $\text{mdxa to+ixu to+ixu onia onia onia onia}$. This is now the equivalent unordered sentence, so we can order the phonemes however we want.

Now, let's say we have scrambled the phonemes in unordered sentence to get

$\text{it+oiunadooaixiaixnotoiaxa+unomn}$.

To perform the ordering transform, we count its phoneme counts, which are, in decimal and binary:

Phoneme	Decimal	Binary
m	1	001
n	4	100
t	2	010
d	1	001
x	1	001
χ	2	010
ɬ	2	010
i	6	110
o	6	110
u	2	010
a	5	101

For the first word — the word with dupe count $k = 0$ — we look at the 1s digit (rightmost digit) of the binary phoneme counts. There are 4 phonemes with a 1 in the 1s digit — m, d, x, a — so the first ordered word is composed of the phonemes mdxa . From the 2s digit, we get that the second word, with dupe count $k = 1$, is composed of tx+iou , and from the 4s digit, we get that the third word, the word with dupe count $k = 2$, is composed of nioa . Therefore, the inverse transform maps the unordered sentence to the ordered sentence mdxa tx+iou nioa . Finally, since words in the ordered language are never differentiated by phoneme order, mdxa tx+iou nioa is equivalent to our ordered sentence at the start, mxad to+ixu onia .

3 Background

The Slau ([ʎau]) are an Austronesian-speaking people native to the Slau islands. There are two main islands, one located in the Devil's Sea in the Pacific Ocean, and the other located in the Bermuda Triangle in the Atlantic Ocean. They migrated from the island of Formosa and settled on the two islands around 2000 years ago.

In the waters surrounding the two main islands are the two openings of the spatio-temporal anomaly known as the Slau wormhole, or the Devil's wormhole, which the Slau use to travel between the two islands. The presence of the wormhole leads to a range of physical and psychological effects in its vicinity, including: abnormal weather, large changes in the magnetic field, anomalous spacetime curvature, disorientation, sneezing, strong winds, left-right confusion, adding irrelevant items to lists, procrastination, and most importantly, time permutation.

With time permutation, segments of time are swapped with each other, and the effect is stronger the closer one is to the wormhole, while it is generally absent on the islands due to shielding from the land mass. This makes communication out at sea difficult, as . Thus on land, they speak the (ordered) Slau language, but out at sea, they speak an unordered language: **Lĩhusã**.

4 Phonology

Non-bold phonemes are “grammatical phonemes”, and **bold** phonemes are “content phonemes”. Grammatical phonemes only occur in affixes, and content phonemes only appear in roots. There are 30 consonants (15 grammatical + 15 content) and 10 vowels (7 grammatical + 3 content)

4.1 Consonants

	Labial	Dental	Retroflex	Palatal	Velar	Uvular	Glottal
Nasal	m	n		ɲ	ŋ		
Plosive	p b	t d	ʈ ɖ	c	k	q	(ʔ)
Sib. fricative		s z	ʂ				
Non-sib. fricative	ɸ β	θ ð		ç	x ɣ	χ	h
Lat. fricative		ɬ ɮ					
Lat. approximant		ɭ	ɮ̥	ʎ			

- /ŋ/: [g~ŋ] word initially and [ŋ] otherwise
- /d/: [d~ɾ]
- /ɭ/: [ɭ~ɮ]
- [ʔ]: non-phonemic prop consonant that is optionally (and often) inserted in vowel clusters to make them easier to pronounce and differentiate

4.2 Vowels

	Front	Central	Back
Close	i ĭ		u ũ
Mid	e ě	(ə)	o õ
Open		a ã	

- [ə]: non-phonemic prop vowel that is optionally inserted in consonant clusters to make them easier to pronounce and differentiate
- “a” and “a” are the same phoneme; this typeface has an unfortunate italics font :(

4.3 Syllable structure

Syllables in Lĩhusã are of the shape CV(C). The prop vowel [ə] and the prop consonant [ʔ] can be inserted to make words conform to the syllable structure. Well, for the listener this doesn’t really matter since the order of the phonemes doesn’t mean anything anyway.

5 Reduplication

Lĩhusã uses the method explained in Section 2 to achieve having free phoneme order. While reduplication is mainly used to replace word order in Lĩhusã, it also has a few other uses that resemble the use of reduplication in natural languages. The former is called *syntactic reduplication*, and the latter *morphological reduplication*.

5.1 Syntactic reduplication

Syntactic reduplication defines the default dupe count of each word in a Lĩhusã sentence. The default dupe count of Lĩhusã is entirely “head least dupe” (which corresponds to head initial word order in ordered languages). In a sentence, the default dupe count is:

predicate < genitive < oblique < nominative < locative < adverb

Nouns have less dupe than the adjectives, genitives, and relative clauses that modify them by default.

5.2 Morphological reduplication

On top of syntactic reduplication, each word can be morphologically reduplicated 0 or 1 times. The effect of morphological reduplication on a word depends on the word class:

For nouns, morphological reduplication marks plurality, and plurality is mandatorily marked by morphological reduplication.

m-xad-0 *to-tixu-1* *o-ŋia-3*
NPST.AV-to.eat-0 OBL.C-fish-1 NOM.C-cat-3
'The cats eat the fish.'

For adjectives, morphological reduplication marks intensity.

ŋia-0 *õ-χtŋl-2*
cat-0 LNK-black-2
'very black cat'

For verbs, morphological reduplication marks intensity or imperfective aspect.

m-xad-1 *o-ŋia-2*
NPST.AV-to.eat-1 NOM.C-cat-2
'The cats eat a lot./The cats devour./The cats are eating./The cats eat (habitually).'

5.3 Deduplication

A problem is that applying morphological reduplication can make a word's dupe count collide with the dupe count of another word, which would violate one of the requirements for the language to be unordered. This is usually fixed by *deduplicating* the word that it collides with, i.e. reducing its dupe count. The procedure is to first perform reduplication, both syntactic and morphological, and then deduplicate words with no morphological reduplication until there are no clashes in dupe count; one may have to deduplicate a word multiple times. This is always possible.

For example, 'black cats' would be

ŋia-1 *õ-χtŋl-0*
cat-1 LNK-black-0
'black cats'

where *ŋia-1* 'cat' has been reduplicated from *ŋia-0*, while *õ-χtŋl-0* 'black' has been deduplicated from *õ-χtŋl-1*, since the two words would otherwise have the same dupe count.

6 Morphology

In Lĩhusã, sentences are made up of *words*, and words are made by combining *roots* with *affixes*. Affixes include *particles*, *voice markers*, and *semantic affixes*.

6.1 Roots

- heavy use of affixes to derive words
- most affixes are productive
- often uses causative affix to form transitive verbs from intransitive verbs
- verbs of emotion/feeling usually have the experiencer of the emotion as NOM and the stimulus as OBL

Lĩhusã root classes are as follows:

- noun (n)
- non-auxiliary verb (v)
- auxiliary verb (vaux)
- adjective (adj)
- adverb (adv)

Note that adverbs are a closed class; usually serial predicate construction is used to modify a verb.

6.2 Affixes

6.2.1 Linking particle and compound predicates

The linking particle (LNK) *õ* (also known as a “ligature”) is used to link multiple predicates together, a noun with a word modifying it, etc.. It is affixed on the modifier.

Noun modified by adjective:

ɲia-0 õ-χɬɲl-1
cat LNK-black
'black cat'

Noun modified by verb:

ɬau-0 õ-m-laxu-1
person LNK-NPST.AV-to.walk
'walking person'

Verb modified by verb (serial predicate): see Section 7.2

6.2.2 Case particles and cases

The other particles mark case. There are 4 cases in Lĩhusã:

	Nominative (NOM)	Oblique (OBL)	Genitive (GEN)	Locative (LOC)
Common (c)	o	to	no	ẽ
Personal (p)	e	te	ne	ẽ

Personal particles are used when the noun is a pronoun, a person's name, refers to a specific person/people, or common nouns referring to people that one wishes to show respect to. Otherwise, common particles are used. Some examples:

e-xu-0
NOM.P-1SG-0
'I'

o-βai-0
 NOM.C-woman-0
 'the woman'

e-βai-0
 NOM.P-woman-0
 'the woman (respectful)'

o-xdi-0
 NOM.C-mother-0
 'the mother'

e-xdi-0
 NOM.P-mother-0
 'mom/my mother'

6.2.3 Voice markers and Austronesian alignment

Lĩhusã, like most if not all Formosan languages, has Austronesian alignment. How this works in Lĩhusã is that there is a system of affixes, called “voice markers”, that have two functions:

- form noun words of various thematic roles (e.g. agent noun, patient noun, etc.) from verb roots (and also convey tense and mood).
- form verb words, where the thematic role of the nominative-marked noun is determined by the affix.

For most sentences, both interpretations are valid and result in the same meaning. The voice markers are as follows:

	Actor voice (AV)	Goal voice (GV)	Locative voice (LV)	Benefactive voice (BV)
Non-past (NPST)	(m)	ũ	ã	ç
Past (PST)	(m)ĩ	ĩũ	ĩã	çĩ
Imperative (IMP)	∅	ĩ	—	—

Note that there exist semantic affixes that look like they are made up of multiple voice markers — these are not considered voice markers; see Section 6.2.4.

We will illustrate the Austronesian alignment system in Lĩhusã with the verb root *βtix* ‘to write’.

Actor voice (AV): Forms the agent from a verb root. *m* is the present agentive affix, so *mβtix* is the present AV form of *βtix*, literally ‘the writer’, ‘the one who writes’. Equivalently, it marks that the nominative noun is the agent.

m-βtix-0 *e-xu-1* *to-bux-2*
 NPST.AV-to.write-0 NOM.P-1SG-1 OBL.C-book-2
 'I write a book'/'I am the writer of a book'

For a small number of verb roots (mostly common intransitive verbs) and semantic affixes, the present and past AV affixes are ∅ and *ĩ* instead of *m* and *mĩ* respectively — the *m* is removed. These are called “m-less verb roots”, in contrast with “non-m-less verb roots”.

Goal voice (GV): Forms the patient or the goal from a verb root. *ũ* is the present patientive affix, so *ũβtix* is the present gv form of *βtix*, literally ‘the thing that is written’ or ‘the thing that results from writing’. As such, the word *ũβtix* means ‘text’ in general. Equivalently, it marks that the nominative noun is the patient or goal.

ũ-βtix-0 *ne-xu-1* *o-bux-2*
 NPST.GV-to.write-0 GEN.P-1SG-1 NOM.C-book-2

'The book is being written by me'/'A book is my thing that is written'

Locative voice (LV): Forms the location or time from a verb root. *ã* is the present locative affix, so *ãβtix* is the present LV form of *βtix*, literally 'the place the writing is done' or 'the time when writing is done'. Equivalently, it marks that the nominative noun is the location or time.

ã-βtix-0 *ne-xu-1* *to-bux-2* *o-γubaχ-3*
NPST.LV-to.write-0 GEN.P-1SG-1 OBL.C-book-2 NOM.C-house-3

'The house is where I write the book'/'The house is my place of writing the book'

Benefactive voice (BV): Forms the beneficiary, recipient, or instrument from a verb root. *ç* is the present benefactive affix, so *çβtix* is the present BV form of *βtix*, literally 'the one the writing is for' or 'the thing the writing is done with'. Equivalently, it marks that the nominative noun is the beneficiary, recipient, or instrument. If the noun is an item, usually the instrumental meaning is meant, while if the noun is a person, usually the beneficiary/recipient meaning is meant.

ç-βtix-0 *ne-xu-1* *to-bux-2* *e-la-3*
NPST.BV-to.write-0 GEN.P-1SG-1 OBL.C-book-2 NOM.P-3PL-3

'They are who I write the book for'/'They are the ones my writing of the book is for'

Imperatives The nominative marks the person being commanded. If this is the second person, it is omitted.

Verbs only have imperatives for the active and goal voices, which are marked by *Ø* and *ĩ* respectively.

Ø-xad-0
IMP.AV-to.eat-0
'Eat!'

ĩ-xad-0
IMP.GV-to.eat-0
'Be eaten!'

Adjectives can also be affixed by *ĩ* to form the imperative.

ĩ-ηuaχ-0
IMP-good-0
'Be good!'

6.2.4 Semantic affixes

- /p/: causative (CAUS) (v -> v.mless), to collect/harvest/produce (n -> v)
- /ɲ/: reflexive (REFL) (v -> v.mless) (shows that agent and patient are the same, or are parts of the same whole)
- /d/: "respectively" (v -> v)
- /dɲ/: reciprocal (RECP) (v -> v.mless)
- /s/: -like (n -> adj), to act/seem like (adj -> v), away (v -> v), go (away) from (n -> v)
- /l/: go to/move towards (n -> v), personal counter
- /k/: come from (n -> v), to do times (numeral -> v)
- /z/: become (n -> v), want to (v -> v)
- /c/: comparative (CMPR) (adj -> adj) ("<NOM> is more than <OBL>")

c-ðluŋ-0 *to-βahu-1* *e-xu-2*
CMPR-big-0 OBL.C-dog-1 NOM.P-1SG-2
'I am bigger than the dog.'

- /t͡ʃã/: superlative (adj -> adj)
- /t/: and (used on case-marked words that syntactically occupy the same position)

- /q/: question marker
- /ĩã/: pertaining to [e.g. for deriving language names from names of groups of people]
- /çã/: manner (v -> v.mless)

ḏaχu-0 çã-βłix-1 ne-xu-2 to-bux-3
 fast-0 manner-to.write-1 GEN.P-1SG-2 OBL.C-book-3
 'I write the book quickly!/'Fast is my writing of the book.'

6.3 Numerals

As Lĩhusã is an unordered language, there are no ordinal numbers. There are, however, cardinal numbers.

7 Sentences

7.1 Elision

Lĩhusã sentences get very long very quickly (exponentially) thanks to the large amount of reduplication. Because of this, words are often dropped if the meaning is understandable from context. That is to say, verb arguments are often dropped. However, the predicate is almost always present.

7.2 Serial predicates

Multiple predicates can be linked together using the linking particle *õ*. When linking multiple predicates together, the nominative marked noun is shared between the predicates. If we have predicates P1 and P2, where P2 has higher dupe count, then linking them would mean ‘P1 and/or then/in order to P2’.

When linking multiple predicates to describe movement, as a rule of thumb, the predicate conveying the *manner* has lower dupe order than the predicate conveying the *direction*. For example, to say ‘I run to the mountain’, in Lĩhusa this would be expressed with

m-βilað-0 õ-m-l-βðd-1 e-xu-2
NPST.AV-to.run-0 LNK-NPST.AV-to.go.to-mountain-1 NOM.P-1SG-2
‘I run to the mountain.’

instead of

m-l-βðd-0 õ-m-βilað-1 e-xu-2
NPST.AV-to.go.to-mountain-0 LNK-NPST.AV-to.run-1 NOM.P-1SG-2
‘I go to the mountain to run.’

7.3 Negation & Prohibitive

Negative sentences are constructed using the negative auxiliary verb (NEG) *di*, which has dupe count 0.

di-0 õ-mĩ-xad-1 to-tiχu-2 e-xu-3
NEG-0 LNK-PST.AV-to.eat-1 OBL.C-fish-2 NOM.P-1SG-3
‘I did not eat the fish.’

The prohibitive (negative imperative) is constructed using the prohibitive auxiliary verb (PROH) *bia*.

bia-0 õ-m-xad-1 to-tiχu-2
PROH-0 LNK-NPST.AV-to.eat-1 OBL.C-fish-2
‘Do not eat the fish.’

7.4 Existence

The existence construction is done with the existential adjective (EXIST) *ẽsa*. This comes from the locative phrase *ẽ* ‘Loc’ + *sa* ‘there’.

ẽsa-0 o-tiχu-1
EXIST-0 NOM.C-fish-1
‘There is a fish.’

Its negative is the negative existential adjective (EX.NEG) *ds*.

ds-0 o-tiχu-1
EX.NEG-0 NOM.C-fish-1
‘There is no fish.’

7.5 Tense

The past/non-past distinction is marked on the voice marker (see Section 6.2.3). The adverb *lu*, when used in conjunction with the non-past voice marker, indicates the future tense (FUT).

8 Writing system

Since the phonemes of each sentence can be ordered any way you want, the Slau take advantage of this with the unordered writing system for Lĩhusã. Also, because it's only used out at sea, Lĩhusã isn't usually written down. When it is written down, it's mainly for ornamental or artistic purposes. Because of this, Lĩhusã writing has evolved to become a blend between art and phonemic representation. There is a lot of freedom and variation in the writing style; here only one style of writing is presented.

The basic principle is that each sentence is a picture with added annotations. The annotations are discrete symbols called *characters*. The placement, rotation, and scaling of the characters in the sentence has no effect on the meaning of the sentence. No two characters are distinguished only by rotation or reflection.

8.1 Case markers

Case is marked using shapes on the edge of objects in the picture. There is no distinction between personal and common case markers.

If possible, the shapes are placed on the object corresponding to the noun it marks. If it is clear from the picture what object is being marked, then the rest of the word might not be written (using content characters). As a result, one might see e.g. a word composed of only a nominative marker.

8.2 Voice markers

There is a character for each phoneme used in the voice markers: /m/, /ũ/, /ã/, /ç/, and /ĩ/. Note that the /m/ character is not used if the verb is *m*-less.

8.3 Other affixes

For other affixes, there is a separate character for each semantic affix.

8.4 Roots

Content characters can be used in 3 ways: logographically, diphonemically, and in compounds.

When used logographically, a character represents an entire root.

Diphonemically, the character represents two content phonemes. Note that since Lĩhusã is unordered, the order of the two phonemes is irrelevant, so since there are 18 content phonemes, there are $\binom{18}{2} = 153$ diphonemic combinations. Note also that some diphonemic combinations have multiple characters that can be used to represent them. Additionally, if there is an odd number of content phonemes in a word, like *xad*, then it can be represented using, for example, characters for *da* and *xa*; when there is an overlapping phoneme, we know that only one copy is meant.

Compound characters can be created by combining characters together. For example, the character used logographically to write *xad* 'to eat' is composed of the characters for 'mouth' and 'taro'; the character composed of 'mouth' plus 'tooth' is used to write the root *xał* 'to bite' or 'to sting'. Compound characters can also be used diphonemically.

Note that most content characters can be used logographically, and all content characters can be used diphonemically. Also, some content characters have a lot of variation when written; for example, the character for 'snake' can be stretched to any length and be curled in any direction.

8.5 Reduplication

When writing, one first picks out the characters to write a word, then all the characters are reduplicated the number of times the word is reduplicated.

