lang_2

grammar of a constructed language

Seth Thompson
a.k.a. kilenc

January 18, 2022

Contents

Part I

Introduction

Background 1

1.1 Origins

lang$_2$ is an *a priori* artlang originally conceived to fulfill a speedlang challenge, then modified to handle a relay, then finally molded into its own project. Although I had no clear motivations in mind when beginning it, I found that the confluence of random decisions—the Gleb-generated inventory, challenge stipulations, the early translation choices—made the project into something I really enjoyed.

1.2 Goals

The overarching goal is to create something that is cool to me. The project is meant to be naturalistic, but when naturalism conflicts with aesthetic, aesthetic will be prioritized.

A primary way to grow this language and develop new ideas will be through translation of poetry and scientific journals. I hope that these will push the limits of my syntactical rules while also developing an interesting corpus. I'll also try to use 5moyds and hopefully at some point a journal to expand my ability to speak the language and get an intuitive sense for what constructions it prefers.

As I develop this conlang, my goals will be to explore analytic constructions and syntactic minutiae, write in-depth documentation of how information structure manifests in the language, and produce a robust dictionary and corpus.

1.3 Lore

lang$_2$ is set in a worldbuilding project I create as a hobby. The language is an isolate spoken on the eastern coast of a peninsula, once the language of city-states and now a common language throughout a number of coastal countries. It is heavily influenced by East Cape, the *de jure* language of the peninsula, and other languages of trade. The world the speakers know is analogous to our 1930s.

Overview 2

2.1 Typology

lang$_2$ is largely head-initial.

There are five total word classes in **lang$_2$**. Content words are largely nouns and verbs, with a small closed set of adjectives. Function words are prepositions and particles.

2.2 Syntax

A basic intransitive clause in **lang$_2$** is composed of a subject followed by a predicate.

(1) **Kanyi hec.**

Kanyi *hec*
Kanyi is there

"It's Kanyi."

(2) **Mavan tahęci.**

mavan *tahęci*
the cat stretches

"The cat's stretching."

In a transitive clause, the object usually follows the verb. Any other dependents are also typically after the object, as in (4).

(3) **Nassoin kąsteci kagesan.**

nassoin *kąsteci* *kagesan*
the king leads the army

"The king leads the army."

(4) **Lalan kirays sapan h-tesan.**

lalan *kiray* *sapan* *ah=tesan*
auntie finds the dog on=the beach

"Auntie's looking for the dog on the beach."

2.3 Morphology

TODO:

Need to finalize the morphological forms etc etc. The main thing is figuring out stress patterns; might need to do some dreaded diachronics.

Nouns inflect for number and deixis.

Deixis is perhaps better characterized as nominal evidentiality; see §7.1.

Table 2.1: Inflection of *metka* "bowl"

	Generic	Proximal	Distal
SG	*metka*	*metkan*	*matkó*
PL	*matkozar*	*métkarran*	*metkazár*

Verbs inflect for person and deictic agreement.

Table 2.2: Inflection of *kęstat* "lead"

	Generic	Proximal	Distal
1.AG	*kęstatr*	*kąstetri*	*kęstatŕ*
1	*kęstatrs*	*kęstatrci*	*kęstatŕs*
2	*kęstata*	*kęstatai*	*kęstatá*
3C	*kęstac*	*kąsteci*	*kęstatés*
3N	*kęstacz*	*kąsteczi*	*kęstatóz*
3N:1	*kęstatns*	*kęstatnsi*	*kęstatńs*
REFL	*kęstatak*	*kęstatki*	*kęstaták*

There are also two nonfinite forms, the bare infinitive and the converb.

Table 2.3: Nonfinite forms of *kęstat* "lead"

Bare	Converb
kęstat	*kąstettal*

Part II

Phonology

Segments 3

lang₂ has 26 phonemic segments, comprising 18 consonants and 8 vowels. The segmental inventory is relatively small compared to the cross-linguistic median, but the ratio of consonants to vowel qualities is the same.

There are two major dialects of **lang₂**: the Port Standard dialect and the Frontier Standard dialect. The Port Standard dialect is more conservative, formed from the dialect region around the northern coastal cities, whereas the Frontier Standard dialect is more innovative, formed from the newer dialect region in the inland cities. The segmental representation of **lang₂** phonology is dialect-neutral, but dialectal realizations of phonemes is discussed when notable.

Discounting tense vowels, the C/VQ ratio in **lang₂** is 3.6; Maddieson (2013) gives the cross-linguistic median as 3.5.

3.1 Consonants

lang₂ has 18 consonant phonemes. There are four distinguished places of articulation: *labial*, *plain alveolar*, *sibilant alveolar*, and *dorsal*, and four distinguished methods of articulation: *nasal*, *stop*, *fricative*, and *glide*. Stops can be further divided into tenuis and voiced, although the voiced counterparts are marginal.

Table 3.1: Consonants

	Labial		Alveolar		Sibilant		Dorsal	
Nasal	m	*m*	n	*n*	n͡z	*z*		
Stop	p	*p*	t	*t*	t͡s	*c*	k	*k*
	b	*b*	d	*d*	d͡z	*j*	g	*g*
Fricative	f	*f*	ɬ	*l*	s	*s*	x	*h*
Glide	w	*v*	ɹ	*r*			j	*y*

Nasal sibilant The nasal sibilant /n͡z/ is a typologically odd segment. The prototypical representation of this sound is [z̃], a voiced fricative with simultaneous nasal release; however, phonetic research suggests that true fricatives cannot be fully nasalized. Thus, this segment's typical realization may be better represented as a slightly fricated approximant [ⁿz̞̃], although some speakers may realize it as a weakly nasalized [ⁿz] or a fully nasalized approximant [ⁿɹ̃]. Nevertheless it patterns as a nasal sibilant in distribution and morphophonemic processes and is thus phonemically represented as /n͡z/.

See Ohala et al. (1998) for further discussion of nasal fricatives.

Labial fricative The fricative /f/ has a marginal distribution, mostly appearing in loan words, names, and onomatopoeia. Most instances of historical /f/ underwent debuccalization and were subsequently lost. Although /x/ also appears frequently in onomatopoeia, it has a more widespread distribution and is not considered marginal.

Sibilant affricate Although the phone [t͡s] appears commonly in the corpus, its underlying form is not always the phoneme /t͡s/. The underlying phoneme is usually elucidated by inflection. For example, both *tsekla* "steering wheel" and *cekla* "great uncle" share the same surface form, [ˈt͡sekɬə]. However, when inflected for the plural, the former becomes [ˌtasəkˈɬažəɹ] and the latter [t͡səkˈɬažəɹ]. As such *tsekla* is analyzed as /tasekɬa/, whereas *cekla* is analyzed as /t͡sekɬa/.

Affrication of stop-fricative clusters and schwa deletion are common processes that yield [t͡s], discussed further in §5.1.2 and §5.2.2.

Voiced stops Voiced stops are often the result of nasal clusters. They are often realized with slight prenasalization, especially in stressed onsets.

Voicing is discussed further in §5.1.1.

Lateral fricative Canonically represented by /ɬ/, the lateral fricative is rather diverse in its crossdialectal realizations. Historically, **l* fricated to dissimilate from **r*, which had become [ɹ] due to its own dissimilation from **z*. In Port Standard /ɬ/ is still realized as [l] except in stressed onsets. In many frontier dialects, it shifted further to [ʃ].

3.2 Vowels

lang₂ has 8 vowel phonemes, 2 high and 6 low. Low vowels are additionally split into tense and lax vowels.

Table 3.2: Vowels

	Front		Mid		Back	
High	i	*i*			u	*u*
Low	e	*ę*	a	*ą*	o	*ǫ*
	ɛ	*e*	ə	*a*	ɔ	*o*

Vowel neutralization Low vowels are reduced to schwa in unstressed syllables. Schwa is not phonemic, but neutralization is common, so it appears frequently throughout the corpus. In speech, the underlying vowel becomes evident when stress is shifted due to morphological processes.

Neutralization is discussed further in §5.2.1.

TODO:

Currently all schwa are romanized as *a* regardless of underlying vowel. Do I want this process to be unmarked instead? That might cause confusion if I ever do some irregular *secondary* stress, but maybe that confusion's good.

Tense vowels Low tense vowels historically derive from nasalized segments, and are still realized with nasalization in the Port Standard dialect. On the other hand, the Frontier Standard dialect typically breaks these vowels, realizing them as falling diphthongs. Because of the disparity of cross-dialect pronunciations, the primary distinguishing feature is tenseness.

Many Frontier dialects further merge /a/ and /o/.

Tense vowels are mostly found in open syllables. Lax vowels are usually found in closed syllables.

3.3 Allophony

4 Supersegments

4.1 Stress

Stress in **lang$_2$** is lexically and morphologically productive. Stress typically falls on the penultimate syllable of a word. Atypical vowel stress, most common in loan words, is marked with an acute. Stress is also morphologically productive, distinguishing between unmarked and distal deixis in nouns. Affixation, compounding, and other stress-shifting processes often cause stress-based minimal pairs to become homophones.

For example, *tąka* “tree” is predictable and unmarked, but *tąká* “reindeer” has atypical ultimate stress and is thus marked.

Stressed vowels have three phonetic differences from unstressed vowels. First, they typically have a rising pitch. Second, they are typically longer than unstressed vowels. Third, onsets before long vowels have a longer VOT than other onsets, a manifestation of slight aspiration or breathiness.

Secondary stress falls on alternating syllables starting from primary stress and spreading left. For example, *kagęsa* “army” has regular stress on the penultimate syllable, but when inflected in plural form, it surfaces as *kegąsazar* [ˌke.gə̃ˈsa.ž̃əɹ]. Secondary stress prevents the reduction of /e a o/ to [ə], but, unlike primary stress, does not cause length or VOT increase.

Morphophonology 5

5.1 Consonants

5.1.1 Voicing

Across morpheme boundaries, clusters composed of a nasal and stop are realized as voiced stops, occasionally prenasalized. Clusters where the nasal is the onset of a syllable, not a coda, do not assimilate, so //pn// is still realized /pn/.

These clusters assimilate in place to the stop, so //np// surfaces as /b/, not /d/.

The assimilation process is the source of all native words with voiced plosives, so the phonemes have a limited and predictable distribution. Foreign words with voiced plosives are often loaned with epenthetic vowels that break up clusters, making the words more closely match native phonotactics.

As such, voiced stops are only marginally phonemic.

Some conjugations of *m*-stem verbs are spelled with a word-final voiced plosive, but the stop is still typically realized as a medial cluster. For example, *seg* "tell themselves" is underlying //sem+k// and realized as [se$^{(n)}$gᵊ].

Some younger speakers always elide the final schwa, especially in the south.

5.1.2 Affrication and assibilation

Across morpheme boundaries, clusters of /t/ and a fricative are realized as a sibilant affricate. Clusters of //t+s//, //t+f//, //t+ɬ// and //t+x// all neutralize to /t͡s/, romanized as *c*.

Clusters of /t/ and the other sibilant phonemes also affricate, but instead of neutralizing to /t͡s/, they have other realizations. Clusters of //t+n͡z// are realized as [t͡ʂ], romanized as *cz*. Clusters of //t+ts// are realized as [t͡sː], romanized as *cc*. In practice, however, these realizations are rare, even when speaking the standardized dialect; most speakers simply render both clusters as [t͡s].

[t͡ʂ] is notoriously hard for non-native speakers to pronounce and is often used as a shibboleth.

Similar to the affrication process, clusters of sibilants and non-sibilant fricative clusters also neutralize. Unvoiced sibilants //t͡s s// clustering with //s f x ɬ// become /s/, romanized as *cc* or *ss* depending on the underlying phoneme. The voiced sibilant //n͡z// becomes /n͡z/ instead, romanized as *zz*.

Sporadically realized as a long [sː] or [ʐ̟ː].

5.2 Vowels

5.2.1 Low vowel neutralization

Low vowels are reduced to [ə] in unstressed syllables. Primary or secondary stress can prevent reduction. The high vowels /i u/ never reduce to schwa.

5.2.2 Schwa deletion

The reduced vowel [ə] is often deleted between consonants, especially between plosives and sibilants. For example, *ksarat* /kosaɹat/ is usually [ksaɹət], and spelled accordingly. When clustering in this way, both /͡ʃs/ and /s/ become /s/. The elision process results in word-final or word-initial [s] or [ǯ] being the only common syllable-internal clusters. However, these clusters are not consistently realized, and occasionally have an epenthetic schwa, especially word-finally.

Nasalized schwa [ə̃] rarely undergoes deletion as it is typically longer than [ə].

TODO:

There might end up being an isogloss map of schwa deletion: most dialects delete ahead of sibilants, some ahead of sibilants and approximants, and others in all environments?

6 Phonotactics

6.1 Roots

Modern native lemma derive from CV or CV(C)CV roots. Some roots have null initials, but they are comparatively rare. Roots allow CC clusters that aren't allowed across morpheme boundaries.

Roots are nominal by default, and in the corpus many roots appear in bare form as nouns. Stems reflect reduced derivational morphemes that transformed roots into verbs or other nouns.

Stem endings can often be reconstructed with vague meanings.

The most prominent verb stem is *-t*, forming both transitive and intransitive verbs with a variety of meanings. Another common stem is *-m*, which also derives stems of either valency class and has a vaguely frequentative connotation. The other rarer verb stems are the intransitive *-k* and the transitive *-l* and *-y*, which encoded some kind of telicity distinction.

Outside of bare roots, *-s* stem is the most common for nouns. There are also *-c* stems and *-z* stems, which likely derived collective and agentive nouns, respectively. All nouns belonging to the sibilant stems are neuter gender. The non-sibilant *-r* stem can form nouns of either gender.

Function words A smaller set of phonemic segments are allowed in functional morphemes. Neither the labials /p f m w/ nor the high vowel /u/ appear in affixes, particles, or prepositions. In the fossilized partial reduplication process, the reflexes of labial consonants are either /k/ (← /p f/) or /n/ (← /m w/). The high vowel /u/ lowers to /o/, often simply realized as [ə].

Because /u/ and /w/ pattern similarly, some phonemic analyses conflate them.

Part III

Morphosyntax

7 Nouns

The **lang**$_2$ noun phrase is largely analytic, but nouns do inflect for deixis and number.

7.1 Deixis

Nominal deixis has a variety of uses, including evidentiality, distance, familiarity, and topicality. Verbs and adjectives exhibit agreement for deictic reference. There are three deictic categories: *generic*, *proximal*, and *distal*.

I had this idea, then found out that, as usual, a natural language had it first. Read Huijsmans, Reisinger, and Matthewson (2020) for more about the Salishan languages.

Although diachronically related to demonstratives, spacial deixis is only a secondary use of the proximal and distal forms. Instead, they are primarily indicators of non-propositional evidentiality, *i.e.* the speaker's evidence of a nominal referent. Evidence in this sense is direct or indirect, the latter largely encoding hearsay or inference. The non-propositional evidentiality system in **lang**$_2$ is semantically based on the speaker's knowledge at or prior to speech time.

Bhadra (2020) terms this a Type I system.

7.1.1 Generic

The generic form of the noun is morphologically least marked and used as the citation form.

The generic form is most often used for gnomic statements.

(5) **Kęsa esyi oc.**
kęsa esyi oc
hero is good
"Heroes are good."

If direct or indirect evidence exists, it's infelicitous to use the generic form.

7.1.2 Proximal

The proximal form of a noun is used when the speaker is certain, nearby, or familiar with the noun. This form most commonly denotes direct evidence, meaning the speaker has personal experience with the marked noun. It is marked with the suffix *-n*.

-n is morphophonemically /⁄on⁄⁄, where /⁄o⁄⁄ doesn't surface for vocalic stems.

Direct evidence The canonical meaning of the proximal form is direct evidence, often translated as "I saw."

7.1.3 Distal

The distal form of a noun is used when the speaker is uncertain, far, or unfamiliar with the noun. This form typically denotes indirect evidence, including inference, meaning the speaker has heard of or can make an educated guess about the existence of the marked noun. Reported deixis is marked by shifting stress to the ultimate syllable of the word.

Indirect Evidence The prototypical meaning of the distal form is indirect evidence, often translated as "heard about" or "they said." As in (6), this evidence is encoded into the clause via the subject and the predicate that agrees with it.

(6) **egi Matkó aczé samés lar t-het.**

egi	*matkó*	*aczé*	*sem-és*	*lar*	*tę=het*
just	basket\DIST	two\DIST	will-3.DIST	there	to=be at

"She said there'll be just two baskets." (5MOYD #1314)

7.2 Number

Nouns inflect morphologically for an additive plural, but there is also a syntactic construction used to form associative plurals. The unmarked form of a noun encodes expected number, *e.g.* *parsa* "eyes" which defaults to a pair and must be modified by a numeral or appositive to specify a singulative.

(a) *matkozar* **(b)** *ezzu metka*

Figure 7.1: Additive vs. associative plurals

The primary difference in the two plurals is the composition of the set: additive plurals refer to largely homogenous referents, whereas associative plurals refer to largely heterogenous referents.

7.2.1 Additive plurals

Additive plurals are used for a set of homogenous referents and never heterogenous referents; *e.g.* *matkozar* is "a set of the same (or similar) bowl" and never "a set of diverse bowls."

The second meaning would use the associative plural.

Additive plurality is indicated through the suffix *-zar*. However, marking is optional and a noun can be inferred additively plural from context. As such marking is less common for small, discrete or easily countable sets or when a referent has been established plural in prior conversation. However, speakers are not always consistent with marking.

Mandatory plural marking is stylistically preferred in formal contexts.

Morphonologically, plural marking can be thought to precede deictic marking; plural distal nouns have accent placed on the plural suffix.

(7) a. **matkó**
"some basket"

b. **matkozar**
"baskets"

c. *__matkózar__

d. **metkazár**
"some baskets"

7.2.2 Other plural strategies

Because the suffixes of *s*-stems and *r*-stems merge in the plural, some minimal pairs are rendered homophones when inflected. There are a few strategies to combat homophony.

Speakers sometimes employ *tevi* "many" as an appositive modifier for the singular form, as in (8a). When number is specified with a numeral, the noun is not marked for plurality, as in (8b).

(8) a. **vęci tevi**
"some mercenaries"

b. **vęci ocza**
"two mercenariesF"

7.2.3 Associative plurals

Unlike the additive plural, the associative plural is not marked with an affix; it is nonetheless rather prevalent. The preposition *ezzu* conveys the associative meaning.

The associative is used for a set of heterogenous referents. For animate (especially human) referents, the meaning is canonically "a person and their associates," as in (9). The focal referent (*i.e.* most important) is the marked noun.

Terminology in this section adapted from Daniel and Moravcsik (2007).

(9) **Sayanalal oti ezzu Kanyi z-laran.**

sayenal-al *ot-i* *ezzu* *Kanyi* *ez=laran*
be ignorant-CVB be-PROX ASSOC NAME in=this

"For this, Kanyi and his friends won't be much help."

However, the associative can also have a number of idiomatic, context-specific meanings, usually referring to diverse groups.

(10) **ezzu Mázzirran segi s-yiat saska hakra.**

ezzu *mázzirran* *segi* *su=yiat* *saska hakra*
ASSOC ballers want:REFL to=steal comeback

"Each and every player wants to pull of the upset."

7.3 Gender

Although nouns traditionally distinguish *common* and *neuter* gender, this has largely become a prescriptive convention. Loanwords, especially technical loanwords, are typically assigned neuter gender. Some words only distinguish gender for certain uses or contexts, thus dictionaries typically denote if a given usage is expected to require neuter gender.

The gender distinction is more common in literature or academia.

7.4 Apposition

Apposition is fairly common due to the small adjective class present in **lang$_2$**. Apposition is used in a number of collocative constructions but is distinct from compounding largely because of stress patterns and morphophonological processes. Furthermore, appositive nouns shown agreement with the noun they modify, like adjectives.

Compounds can shift stress and also show cross-morpheme sound changes.

7.4.1 Inalienable possession

Apposition is a fairly common strategy for possession of all types, but inalienably possessed nouns require apposition, as in (11a).

Alienable possession via apposition is dispreferred in formal registers.

(11) a. **hora kęsa**
"soldier's wrist"

b. ?**hora im-kęsa**
"soldier's wrist"

The prepositional construction more readily lends itself to a reading indicating inalienable possession, such as "soldier's craftmanship" or "soldier's weaponsmithing."

7.5 Phrasal syntax

Noun phrases are predominantly head-initial. Generally speaking, syntactically simpler constituents occur before more syntactically complex ones.

(12) head → adjective → number → appositive → prepositional phrase

8 Pronouns

Pronouns are morphologically and syntactically similar to other nouns—they share inflectional patterns and can be modified by adjectives. The notable differences are that pronouns rarely inflect for deixis, and some appositive constructions are ungrammatical.

TODO:

Settle on pronominal forms—right now it's *rat* 1, *a(f)* 2, *sec* 3C, and *moc* 3N. There's some isogloss map about whether 2 is *a* or *af*.

Possession Like nominal possession, pronominal possession can be expressed either through apposition or prepositional phrases. Thus both *lar im-sec* and *lar sec* can mean "his thing." However, apposition is more common in isolation for pronouns than it is for nouns.

Formal registers prefer the prepositional construction.

9 Verbs

Like the noun phrase, the **lang₂** verb phrase is mostly analytic, and inflection is largely reserved for agreement. Other parts of the verb complex are handled by periphrastic constructions, notably including valency operations and TAM.

9.1 Agreement

Verbs display agreement along two axes: *deixis* and *person*. Deictic agreement is only with the subject of the clause, but person agreement is with both subject and the object, if the verb is transitive. Only the sole finite verb of a clause bears agreement; verbs demoted to adjuncts or arguments are always uninflected.

Despite person agreement, **lang₂** is rarely pro-drop.

9.1.1 Person

In old **lang₂**, verb agreement was transparently derived from cliticized pronouns. However, sound change reduced agreement endings, leading to synchronic forms that are often actually patient agreement.

TODO:

How to cover morphophonological stuff? The two main ones are affrication of /t/ + /s z/ and cluster-final /s ts/ merging to [s], spelled *s*.

Except in a few outlier cases, agreement is fairly predictable from Table 9.1. Neuter agreement takes precedent over other morphemes when applicable. Furthermore, *-a-* only appears on intransitive verbs; for transitive verbs, the appropriate agreement with the other argument is used instead. Finally, *-r-* is used for intransitives, not the agent form.

When attached to a finite verb without deictic agreement, person agreement morphemes are realized as short: unlike many other affixes, they do not shift stress. This realization preserves old stress patterns before the morphemes were reduced. Because of this quirk, some scholars argue that person agreement morphemes are actually clitics to which deictic affixes attach.

Morpheme	Meaning
-rc-	1.AG
-r-	1.PT
-a-	2
-s-	3C
-z-	3N
-ns-	1↔3N
-k-	REFL

Table 9.1: Summary of person agreement

1st-person agreement is ergative as an accident of sound change.

TODO:

Do I want to keep this or say they got analogized? Maybe there's some weird interactions with the stress-fixing rules I've been cooking up?

Because **lang$_2$** allows frequent fronting, a clause's agent and patient can become ambiguous when the agreement morphemes are not sufficient. This especially occurs with common and neuter agreement. A mediopassive construction formed with *het* can be used to clarify such instances. In the mediopassive, the semantic verb is demoted to adjunct as a converb.

The *het* construction is not a true passive: the verb does not change valency (*i. e.* the A-like argument can't be omitted).

(13) **Kanyi akvací Arpat.**

Kanyi akvací Arpat
NAME tagged:3C NAME

"Kanyi tagged Arpat"
Or: "Who Arpat tagged was Kanyi"

(14) **Kanyi hací Arpat akvetal.**

Kanyi hací Arpat akvetal
NAME be:3C NAME tagging

"Kanyi was tagged by Arpat"

In (13), it's not clear if Kanyi is the tagger or the focus-fronted taggee; both interpretations are grammatical. The use of the passive in (14) is less ambiguously interpreted, always meaning that Arpat is the semantic agent.

9.1.2 Deixis

Generic noun forms do not have agreement morphemes, but proximal nouns demand the verbal suffix *-i* and distal nouns shift stress to the final syllable. These suffixes are attached after person agreement suffixes. The proximal suffix *-i* is often stressed itself.

The stress shift is a remnant of an elided morpheme.

TODO:

Same question as earlier: do I like the stress rules for this affix? Or the way this next one preserves things?

Due to historical stress rules, the distal agreement suffix is more fusional for some select verb forms. Although the vowel was lost from the common and neuter agreement suffixes in other forms, it remains in the distal form.

Generic		Distal	
-s	9	*Verbs -és*	22
-z		*-óz*	

Table 9.2: Deictic forms of agreement

Deictic agreement requires verbs to encode the same evidentiality as the subject noun phrase, but this may be infelicitous in some contexts. The *het* construction can also be used to change verbal agreement.

(15) * **Azzár hossusassi nassoin.**

as-zár *hassusal-s-i* *nassoi-n*
foreigner-PL.DIST exalt-3C-PROX king-PROX

Intended: "(I see) the foreigners (I've heard about) praising the king."

(16) **Nassoin heci azzár hossusalal.**

nassoi-n *het-s-i* *as-zár* *hossusalal*
king-PROX be-3C-PROX foreigner-PL.DIST exalting

"(I see) the king being praised by the foreigners (I've heard about)."

In (15), the speaker intends to mark the verb as proximal to convey direct evidence, but the utterance is ungrammatical because the verb doesn't agree with its subject, *azzár*. To correct this, the construction in (16) is used, which takes advantage of the passive to mark the verb phrase as proximal.

9.2 Converb

The converb form of a verb is used for simultaneous action. The converb is commonly used to describe the manner of the main clause, and is also commonly used in periphrastic constructions. The converbial suffix is *-al*.

TODO:

Work more on the converb and flesh this out with examples. Might change the morphophonemic form. I've flip-flopped on this one a billion times so maybe I'll do some different ones for different stems.

9.3 Transitivity

Transitivity is lexically set. There are three valency classes a verb can fall into: *transitive*, *intransitive*, and *pseudo-transitive*. Transitive verbs always have two arguments and intransitive verbs always have one. Pseudo-transitive verbs also take more than one argument, but are morphologically intransitive, and as such their additional argument is a prepositional phrase that cannot be ommitted.

All valency classes allow a number of optional but often collocated adjuncts, introduced as prepositional phrases or converbs.

Non-finite verbs of each type do not have fixed valence and can omit all arguments. In some periphrastic constructions, those arguments are still required by the new finite verb. However, other constructions, and general adjectival or adverbial use, often appear without overt arguments.

There are few methods to decrease a verb's valency, which is usually done when an argument is sufficiently clear from context. In contrast, there are no methods to increase a verb's valency.

The most common method of valency reduction is dummy objects. Most transitive verbs have a collocated dummy object. Prescriptive convention holds that verbs exhibit person agreement with these objects, but in speech these verbs are often treated as morphologically intransitive, as in (17b). Although colloquial, this phenomenon points to further grammaticalization of dummy objects.

Some verbs have multiple collocations for different senses.

(17) a. **a Semsi sora.**

a semsi sora
you say:3C DMY
"You're talking."

b. **a Sémai lar.**

a sémai lar
you say:2 DMY
"You're talking."

Informally, *lar* is more common than collocative dummy nouns.

For verbs that lack a distinct collocated intransitive form, or for certain pragmatic reasons, a periphrastic construction can also serve as a valency-changing operation. The intransitive copula *ot* is the most common, demoting the semantic verb to adjunct as a converb.

(18) **Nassoin oci kąstetal.**

nassoin oci kąstetal
the king is leading
"The king's in charge."

Although rarer than object omission, subject omission is accomplished through the *pit* passive. The passive construction with *pit* promotes the object to subject and demotes the semantic verb to an adjunct with *ez*.

TODO:

Some more info about using the reflexive to reduce valency, too.

9.4 Negation

Negation can be handled in multiple ways. The typical method is a periphrastic construction with the verb *rek*. Other methods include the use of discourse particles ...

TODO:

Write the section about which particles can convey negation, when, and how etc.

9.5 Aspect and mood

9.5.1 Perfective

The perfective construction uses the instransitive copula *ot*, demoting the semantic verb to adjunct with the preposition *tę*. The perfective can be used in any time frame, although by default it does have a past-time connotation.

The perfective construction generally used for events that occured over a fixed time frame, especially when a duration is given. In contrast with an unmarked verb, the perfective emphasizes sequences of events or actions done a finite number of times.

9.5.2 Irrealis

The imperfective construction uses *sem* “say, want,” demoting the semantic verb to adjunct with the preposition *tę*. The irrealis can be used in any time frame, although by default it does have future-time connotation.

The irrealis is used for all events that the speaker supposes might or might’ve occured. The construction is very general, and has broad semantic meaning—including conditional, jussive, and optative senses. In contrast with the unmarked verb, the irrealis emphasizes that the situation is not factual, but is hoped or posited to occur or have occured.

9.5.3 Subjunctive

The subjunctive construction uses *nenat* “crouch,” demoting the semantic verb to adjunct with the preposition *ah*. The subjunctive construction has a more limited scope than the *sem* irrealis construction, typically expressing counterfactuals or doubt.

10 Adpositions

Adpositions are syntactically bound morphemes that express some relationship (often spacial) between constituents. However, they are considered words, not affixes, because the stress pattern of the noun they bind to does not shift.

Compare *kąsazar* "soldiers," marked via affix, to *retus im-kęsa*, "soldier's blade," marked via adposition.

Adpositions are a closed class, composed of only 6 members; finer distinctions can be made with *compound adpositions*, such as *tę-kamc im* "after, to the back of." Although many such constructions are common enough to be lexically set, they are not nearly as ubiquitous as lone prepositions.

TODO:

Should I go with the East Cape spelling conventions? They'd be one letter instead of the full form, eg. *t-kamc*. I'm thinking that even if they're spelled this way they might not even be clitics? They're phonologically independent it seems like.

10.0.1 im

The adposition *im* indicates possession, often alienable. It can also indicate origin.

10.0.2 ez

The adposition *ez* conveys location inside an object or large body. It can also be used for composition of manmade objects.

10.0.3 tę

The adposition *tę* conveys motion relative to a location, either towards or away from.

10.0.4 ah

The adposition *ah* conveys location on the surface of another object. It can also be used for general location.

10.0.5 osc

The adposition *osc* conveys location surrounding another object. It is commonly used in a temporal sense to indicate a time frame, often translated as “around the time of.”

10.0.6 u and su

U and *su* are more limited in semantic scope than other adpositions and are rarely used in compound prepositions to gain further nuance. However, they have similar syntactic distribution and are thus considered members of the preposition class. They typically convey association alongisde.

While *u* is used largely to join two noun phrases as arguments of one head, such as *laran u-lár* “this and that,” *su* is used for emphasize or in some fixed constructions, such as *su-kagęsa su-kagęstapa* “both army and navy.”

11 Adjectives

Adjectives are a small, closed class of noun-like morphemes that cannot be arguments of verbs or prepositions without some other constituent. There are approximately 20 adjectives. Adjectives agree with their head noun for deixis (but not number) and syntactically appear before appositives or prepositional modifiers in the noun phrase.

12 Particles

Particles are a small but open class of discourse markers that can appear at the beginning of a clause. The most common particles are *hes*, a polar question marker, *vi*, a content question marker, and *ǫm*, an imperative marker. They are syntactically privileged, able to occur before other constituents of a clause, including even fronted arguments or adjuncts.

(19) **yiz: Otr. vi: Mans pesayrci rat, vi: Mans kirayrci rat?**

yiz	*ot-r*	*vi*	*mans*	*pesay-rci*	*rat*	*vi*	*mans*	*kiray-rci*	*rat*
then	be-1	Q	who	marry-1.AG	1	Q	who	find-1.AG	1

"Now then—who do I marry, who do I find?" (5MOYD #1518)

Only one particle can appear in a clause.

Clauses 13

Although the base-generated word order in **lang₂** is SVO, this order rarely surfaces due to aggressive focus fronting. The most proximal, most newsworthy information is placed in the front of an utterance in first position. As a consequence, **lang₂** is V2 order, mandating that a finite verb always be in second position. Adjuncts, including demoted verbal constructions, typically come after the core arguments of the verb. Often, however, the order of elements in a clause is determined by focality and evidentiality.

The most common word order in simple declarative sentences is SXOV or VXOS.

13.1 Fronting

The most likely phrases to be fronted are proximal or directly evident noun phrases, followed by distal or indirectly evident noun phrases. Generic noun phrases are rarely fronted except in fixed constructions.

Often, the fronted element will be the conversational focus.

lang₂ doesn't have explicit role-marking, so it's not always clear if a fronted noun phrase is subject or object. Various structures, including verb agreement and passives, help provided redundancy when context is not enough.

Part IV

Appendix

Culture A

A.1 Metaphors

Some common conceptual metaphors in **lang$_2$** are given below; most things are also noted in the dictionary.

WISDOM IS A TREE Trees are an important part of **lang$_2$** culture, representing the emotional and intellectual lifespan of a person. Old, tall deciduous trees are a symbol of widsom and maturity; as a plant may *ossat* "grow," so too may a person *ossat* "become wiser." Young children attend a *pebar* "garden" for primary education. A student might lament that a former mentor *tąvat* "becomes senile (*lit.* loses leaves)," or that an elder has "thinning branches."

POLITICS IS A SHORE The coastline is a common conceptual metaphor for things involving governance. A common idiom for kingdom or state is *taspa u tesa* "sea and shore," its citizens are *adahęs* "sands," its influence *almanizar* "waves." A tribune might *kirǫyam* "investigate (*lit.* delve into)" an issue, or a *cunvarą* "columnist (*lit.* seagull)" may discuss it.

BODY PARTS ARE SKILLS Traditional **lang$_2$** culturual and spiritual beliefs posit that the person is divided into two parts, a perfect body and an imperfect mind. As such many body parts are sacred to certain skills or traits—*hora* "wrist" represents craftmanship; *ahka* "foot" represents wisdom or learnedness; *kamc* "back" represents labor; *makra* "chest" represents responsibility. Someone with poor skill may be *orra* "feeble" or *???* "uncoordinated," while someone unresponsibile may be *oca* "thin."

Registers B

B.1 Poetry

The common structure of a classical $\mathbf{lang}_2$ poem has nine lines, with alternating pairs of short lines and long lines. Short lines are two trochees between two amphibrachs, and long lines are four amphibrachs. The middle three lines of the verse share a central, interrelated metaphor, and the first and last lines feature repitition. Common poetic features also include alliteration across lines and rhyming of stress syllables. Syncope is often used to fit the meter.

Lexicon C

A **lang$_2$**-to-English dictionary is provided below.

How to use

Entries for lexical items are listed by their spelling in generic form, ignoring morphophonological alterations. Derived words are listed as separate entries, but their source word is given. On the other hand, idiomatic or fixed expressions are given under the lexical item.

Each sense of a word has three basic parts: a quick, single-word translation for ease-of-use; a more detailed explanation of the concept; and an example sentence. The sentences are usually designed to help the reader figure the word's meaning out from context, particularly for **lang$_2$** speakers and learners.

Pronunciations are given dictionary-style in a phonetic alphabet more intuitive to native **lang$_2$** speakers.

A

asyisaikli • *noun* • /ěsyisŏíkli/ • **1** organization, association, official body

avąra • *noun* • /avą́rě/ ⟵ OL_2 *agą̆ré*, CC *gamrī* • see *secya* • **1** *ntr.* star or constellation used for navigation

Arméds • *noun* ⟵ Farlands *rmědas* • **1** *ntr.* the Gospels; a collection of poems sacred to the Véajan religion

Ą

ąkassar • *noun* • /ǫ̆kássŏr/ ⟵ OL_2 *ǫ̆kássŏr* "payment craft" • **1** *ntr.* mathematics

C

E

egi • *particle* • **1** just, only: *egi matkó aczé sém lar tę-het* "she says there'll be just two baskets"

emas • *noun* • **1** *ntr.* plot of land

emassoi • *noun* ⟵ *emas* + *-soi* • **1** boss; mid-level employee that oversees other employees **2** coach, manager, trainer; person with authority over a sports teams or its players **3** **emassoi kąstezi** coach; sports executive that makes substitutions and decides strategy **4** **emassoi ǫkas** general manager; sports executive that signs and trades players **5** *archaic* feudal lord

esyi • *adj.* • **1** good **2** correct, appropriate

Ę

H

hassusar • *verb tr.* • /hŏssúsar/ ⟵ OL_2 *hŏs-húsăr*, redup. of *húsăr* • **1** exalt, praise (smn.) **2** *refl.* boast about *tę* smth. **3** be a fan of, root for, cheer for (a team)

hakra • *noun* • **1** battle, skirmish **2** *pl.* military campaign **3** *pl.* semester

hakraif • *noun* ⟵ *hakra* + *-if* • **1** field hockey; a ball game played on a pitch between two teams of seven, each attempting to win by scoring more points via goals than the opponent

husar • *verb tr.* • **1** compliment **2** *archaic* shout at

I

ista • *noun* • /ístŏ/ ⟵ Farlands *ísto*, *hísto* • **1** *(Véajan) ntr.* grace; a quite prayer ceremony for blessing food, travel, and other small events **2** **tahąt osc-ista** give grace **3** **fala ista** chamomile tea; a kind of soothing herbal drink imported from the Farlands

K

kagęsa • *noun* • /kě(n)gę́să/ ⟵ OL_2 *kę̆kę́să*, redup. of *kę̆să* • **1** battalion, unit **2** army; a branch of the military comprising ground troops **3** *sports* team, club

kęstat • *verb tr.* • /kęstět/ • **1** lead (smn.) towards *tę* a goal **2** command: *nassoin kęstací su kagęsa su kagę́staspa* "the king commands both army and navy." **3** train (an apprentice) in *tę* a skill: *tę-coryam kęstatrí rat lala* "auntie's teaching me to sew." **4** formally teach (a student) in *ez* a discipline: *rappahan picí kąstetnal ez-latya* "the minister was brought up in the faith."

kęsa • *noun* • **1** soldier **2** *literary* protagonist, hero

L

M

mazzi • *noun* • **1** ballplayer

N

nahhar • *noun* • /náhhŏr/ ← OL_2 *náhhŏr* "gem craft" • **1** *ntr.* ring, hoop **2** *ntr.* bracelet, bracer **3** *ntr. obs.* jewelry

O

Q

ǫkas • *noun* • **1** *ntr.* payment

P

palác • *noun* • /pălát(ă)s/ • **1** *ntr.* dinner

paknalác • *noun* • /paknălát(ă)s/ ← pormanteau of *pakna u-palác* • **1** inn, motel; short-term lodgings for a traveller between urban areas

R

S

Samvéajan • *noun* ← compounding of *sem véaja* • always PROX • **1** *ntr.* the Gospels; a collection of poems sacred to the Véajan religion

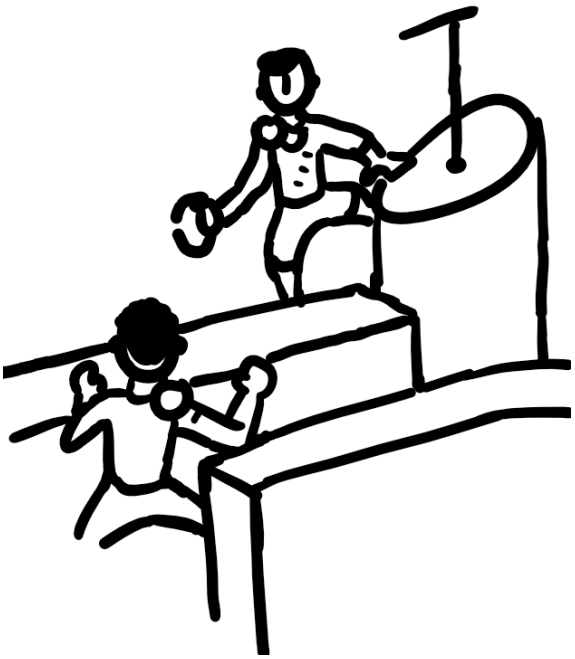
sem • *verb tr.* • **1** say to, talk to **2** *refl.* tell that **3** *refl.* say **4** **sem ???** talk, speak **5** *coll., intr.* talk, speak **6** speak on, give an opinion on smth. concerning *ez* **7** **sem véaja** preach on *ez* **8** **sem ???** give a speech about *ez* **9** **sem ???** lecture on *ez* **10** want smn. to *im* do **11** *refl.* want to *su* do **12** **sem tę** would **13** **sem tę** ought **14** **sem tę** hopefully **15** *(of symbols, texts)* signify **16** *(of words, phrases)* mean, be defined as **17** *(of title, roles)* enable, allow to **18** **sagí tę** *coll., refl.* let's go to *tę*: *sagí tę-tesan* "let's go to the beach" **19** **sagí tę** *coll., refl.* let's *tę* do: *sagí tę-kęstat lala* "let's hire a nanny" **20** *coll.* looks like smn. **21** **samés** *coll., intj.* looks like!

sisik • *noun* ← clipping of *tahęnhar asyisaikli* "associated tag" • see *tahęnhar* • **1** *Inland* tag

T

tahąt • *verb in.* • /táhęt/ • **1** stretch (one's limbs) **2** reach for *tę* something (to grab it) **3** aspire to *tę* a difficult goal: *tę-mazzi im-ot tahącí sec, manacan acoan raksí otnal kers* "he aspired to be a ballplayer, but he wasn't tall enough" **4** pray about *osc* smth.

tahęnhar • *noun* • /tăhęnhŏr/ ⟵ OL_2 *tahębnáhhŏr* "hoop reach" • **1** *ntr.* tag; a parkour game played on an obstacle course between two teams over a number of rounds, each attempting to win by grabbing rings before being tagged



taspa u-tesa • **1** *idiom* kingdom

tesa • *noun* • **1** shoreline, bank; where a moving body of water meets the land **2** *of people* hair line

teukkanas • *noun* • /teukkánŏs/ ⟵ Farlands *tiēkʸkános* • **1** *(Véajan) ntr.* missionary, pastor; travelling evangelist who starts local chapters of faith centers **2** *humorous* tourist, vacationer

U

V

véaja • *noun* • /véă(n)jŏ/ ⟵ Farlands *vḗzṣā* • **1** *ntr.* karma; recompense for virtuous behavior **2** *ntr. pl. prox.* a minority religion characterized by proselytization and belief in an afterlife **3** **sesam véaja** preach, orate: *rappahan sesamzí véaja ez-ǫkas esyi tę-laczin* "the minister preached of an eternal reward for the faithful"

Y

yerragín • *noun* • /yerrŏ(n)gín/ ⟵ OL_2 *yer-vŏngín*, Farlands *iēlunzṣín* • **1** *(Véajan) ntr.* a prayer ceremony for greeting a visitor as a show of hospitality **2** **tahąt osc-yerragín** lead the prayer ceremony **3** *idiom* **yiat yerragín** show off, grandstand, put on airs: *egi cuvarąn yiací yerragín tę-ahka im-nassoin* "the king knew the jester was only showing off"

Z