Solution summary:

- word order: (S) (O) V or (S) V$_1$ V$_2$ (O)

- composition of verb (V$_1$ only has stem and TAM if present):

$$
\left\{
\begin{array}{ll}
\varnothing & \text{affirmative} \\
\textbf{ya} & \text{negative indicative} \\
\textbf{ʔu} & \text{negative imperative}
\end{array}
\right\}
+
\left\{
\begin{array}{ll}
\textbf{-ʔən} & \text{1erg} \\
\textbf{də} & \text{1abs} \\
\textbf{-ʔəm} & \text{2nom} \\
\textbf{mi} & \text{2acc} \\
\textbf{-ʔəy} & \text{3erg} \\
\varnothing & \text{3abs}
\end{array}
\right\}
+
\left\{
\begin{array}{ll}
\textbf{yak-} & \text{causative} \\
\varnothing & \text{otherwise}
\end{array}
\right\}
+ \text{verb-stem} +
$$

$$
\left\{
\begin{array}{ll}
\textbf{-ʔəm-} & \text{passive} \\
\textbf{-ʔoy-} & \text{antipassive} \\
\textbf{-hay-} & \text{applicative} \\
\varnothing & \text{otherwise}
\end{array}
\right\}
+
\left\{
\begin{array}{ll}
\textbf{-tam-} & \text{1st/2nd person plural} \\
\textbf{-šuk-} & \text{3rd person plural} \\
\varnothing & \text{all singular}
\end{array}
\right\}
+
$$

$$
\left\{
\begin{array}{ll}
\textbf{-pa} & \text{affirmative present} \\
\textbf{-A} & \text{negative present} \\
\textbf{-E} & \text{dependent past} \\
\textbf{-Aʔ} & \text{affirmative imperative} \\
\textbf{-ʔə} & \text{otherwise}
\end{array}
\right\}
+
\left\{
\begin{array}{ll}
\textbf{-ʔam} & \text{"now"} \\
\varnothing & \text{otherwise}
\end{array}
\right\}
$$

- split ergativity: ergative behavior for 1st and 3rd person, accusative for 2nd

- direct-inverse alternation: verbs marked by highest actor in hierarchy $1 = 2 > 3$

- vowel harmony: /A/ is [a] after [e, o] and [ə] elsewhere, /E/ is [i] after [i, u] and [e] elsewhere

- phonological rules: tš, cš > č; y disappears between a and e; yT > D (where T is a voiceless stop and D is the corresponding voiced stop); [ŋ] syllable-finally and [w] elsewhere are in complementary distribution, [ʔ] disappears syllable-finally except when utterance-final

- a ʔ-initial suffix causes gemination of a preceding consonant before the ʔ is deleted; if this causes two vowels to become adjacent, one is deleted (preferentially the schwa)

1. *swelling*
2. *You all are beginning to burn (it).*
3. *They are not dancing.*
4. *They are going to eat the meat.*
5. *He/she is making him/her see.*
6. *You didn't walk for him/her.*
7. **puci**
8. **šeme kəššobám**
9. **toto poŋŋəmpa**
10. **šiššən hokəštaməʔ**
11. **gahi ya wičukə**
12. **nəkpan tətammə nəʔ**

A detailed explanation is provided on the following pages.

This problem primarily focuses on various phonological phenomena, but it also involves valency-changing affixes as well as morphological derivations.

Here are glosses of all of the sentences given:

(A)
ʔokmaŋpa   haʔčukkə
ʔokmaŋ-pa=Ø   haʔc-šuk-wə
begin-INC=3.ABS go_into_debt-3PL-DEP.INC
They are beginning to go into debt.

(B)
cəkhae   ʔum   ʔiššə
cək-hay-E   ʔu=ʔəm   ʔiš-wə
do-APPL-NMLZ   neg.IMP=2.NOM see-NEG.IMP
Don't see the witchcraft.

(C)
šiššən   poŋpa
šiš=ʔən   poŋ-pa
meat=1.ERG burn-INC
I am burning the meat.

(D)
nə   yay   močukkə
nəʔ   ya=ʔəy   mot-šuk-wə
water NEG=3.ERG mix-3PL-NEG.COM
They didn't mix the water.

(E)
pən minpay   šemhayyám   yomáʔ
pən   min-pa=ʔəy   šem-hay-wə-ʔam   yomáʔ
man come-INC=3.ERG play-APPL-DEP.INC-NOW woman
The man is coming to play (music) for the woman now.

(F)
ya   mi   yakhacə
ya   mi   yak-haʔc-A
NEG 2.ACC CAUS-go_into_debt-NEG.INC
He/she is not making you go into debt.

(G)
ʔokmawə də   wittammám
ʔokmaŋ-Aʔ   də   wit-tam-wə-ʔam
begin-IMP   1.ABS walk-1/2PL-DEP.IMP-NOW
Let's begin to walk now.

(H)
hopowe   ʔun   cəktammə
ho-poŋ-E   ʔu=ʔən   cək-tam-wə
INTO-burn-NMLZ   NEG.IMP=1.ERG do-1/2PL-NEG.IMP
Let's not make the fire.

(I)
yomá   təʔʔoyyə
yomáʔ=Ø   təʔ-ʔoy-wə
woman=3.ABS want-ANTIP-COM
The woman wanted.

(J)
nəkkən   hoʔištame   totoʔ
nək-wə=ʔən   ho-ʔiš-tam-E   totoʔ
go-COM=1.ERG INTO-see-1/2PL-DEP.COM book
We went to examine the book.

(K)
ʔukkoba   hukkoba
Ø   ʔuk-ʔoy-pa   huk-ʔoy-pa
3.ABS drink-ANTIP-INC smoke-ANTIP-INC
He/she is drinking and smoking.

(L)
yəhi   gadə   cəkkəmmə
yə-hi   ga-də=Ø   cək-ʔəm-wə
PROX-LOC DIST-DEM=3.ABS do-PASS-COM
That was done here.

(M)
ʔəm   ʔectampa
ʔəm   ʔec-tam-pa
2.NOM dance-1/2PL-INC
You all are dancing.

(N)
nə   yan   hea
nəʔ   ya=ʔən   hey-A
water NEG=1.ERG stir-NEG.INC
I am not stirring the water.

(O)
huki   šiŋpám
huk-E=Ø   šiŋ-pa-ʔam
smoke-NMLZ=3.ABS swell-INC-NOW
The cigarette is swelling now.

(P)
yədəm   motaʔ
yə-də=ʔəm   mot-Aʔ
PROX-DEM=2.NOM mix-IMP
Mix this.

(Q)
nəʔʔəm   hoʔukəʔ
nəʔ=ʔəm   ho-ʔuk-Aʔ
water=2.NOM INTO-drink-IMP
Drink all of the water (to the last drop).

(R)
šiš   puʔccəmmə
šis=Ø   puʔc-ʔəm-wə
meat=3.ABS rot-PASS-COM
The meat was rotted.

(S)
də   yakkəššodammə
də   yak-kəš-ʔoy-tam-wə
1.ABS CAUS-eat-ANTIP-1/2PL-COM
He/she fed us.

(T)
minnəy   heyšuki   šiš
min-wə=ʔəy   hey-šuk-E   šiš
come-COM=3.ERG stir-3PL-DEP.COM meat
They came to stir the meat.

I will now outline a potential chain of reasoning by which the language's rules can be deduced.

On the surface, it seems like the sentences share little in common. In particular, only 6 verbs appear to be repeated: *begin, come, drink, go into debt, mix*, and *stir*. Since it is difficult to compare only two sentences, especially when similarities are hard to find (e.g. (A) and (F) look very different despite having the same verb), it may be easier to start by comparing the nouns.

The most notable are *meat* and *water*, which have 3 occurrences each. Starting with the former, we see that sentences (C), (R), and (T) contain the words **šiššən**, **šiš**, and **šiš** respectively. Setting aside the anomalous suffix for now, it seems reasonable to look at (T) in more detail, as it shares two verbs with other sentences. Comparing with (E), which also has *come*, **minpay** and **minnəy** are the only words which appear vaguely similar, so we can hypothesize that **min-** or something close encodes the verb *come*.

Deducing the translation of *stir* is a bit more involved; there is little in common between (N) and (T), and we have no reason to prefer **-nəy** or some part of **heyšuki**. However, we can make a reasonable guess by process of elimination. Looking at (D), (N), and (Q) suggests that **nə**, **nə**, and **nəʔʔəm** respectively describe *water*, which rules out the first word of (N). The tenuous connection between **hea** and **heyšuki** can be reinforced by (helpfully) noting that the negative declarative (non-imperative) sentences, (D), (F), and (N), contain **yay**, **ya**, and **yan**. Thus we have decomposed the content words of sentence (T) with reasonable confidence as something along the lines of *come – stir – meat*, in addition to (N) as *water – not – stir*.

Continuing with this process, we can see that the last word of (D), which we currently believe to be *water – not – ???*, likely encodes the verb, *mix*. The closest parallel in (P) appears to be **motaʔ**, which bears a two-letter prefix in common with **močukkə**. While a tenuous connection, this agrees with our current hypothesis. After observing the obvious similarities between (E) **yomáʔ**∼(I) **yomá** (*woman*) and (A) **ʔokmaŋpa**∼(G) **ʔokmawə** (*begin*), we have exhausted all the immediately evident comparisons.

We can now begin to examine broader and more general patterns. The simplest of these is word order. Among the "simple" sentences—(C), (D), (I), (N), (P), (Q), (R)—we note that the word order is consistently the noun (subject or object), followed by a negation particle, followed by the verb.[1] As for the sentences with two verbs—(A), (E), (G), (T)—we see the verbs in the same order as in English and the object at the end. This is a useful observation because it allows us to identify the individual words in (J) and (O): *go – examine – book* and *cigarette – swell*.

With a reasonable amount of data we're relatively sure about, we can now begin looking at grammatical particles and affixes. First, tense[2] seems like a good place to start, as each of the sentences that we have identified so far (except (G), (P), and (Q), which have some form of imperative mood) can be classified as either past or present. Hence:

	past		present
(D)	**močukkə**	(A)	**ʔokmaŋpa haʔčukkə**
(I)	**təʔʔoyyə**	(C)	**poŋpa**
(J)	**nəkkən hoʔištame**	(E)	**minpay šemhayyám**
(R)	**puʔccəmmə**	(N)	**hea**
(T)	**minnəy heyšuki**	(O)	**šiŋpám**

When sorting the verbs like this, some clear patterns emerge. In the first column, all of the

single verbs end in a long (geminated) consonant followed by ə. In the second column, the single verbs end in **-pa** with the strange exception of **hea**, which we'll keep in mind and go back to later. (**šiŋpám** is also different, but you may have already observed that all the sentences containing the word "now" have verbs ending in **-ám**, and it seems reasonable that this suffix would absorb the final **a** in **-pa**.) The sentences with two verbs follow the same pattern for the first verb, although they sometimes have some strange extra suffix (**nəkkən**, **minnəy**, **minpay**) that we will have to explain.

With tense mostly figured out, we can move on to looking at person. Here are the same ten sentences sorted by subject, with the verbs boxed:

1sg	(C)	šiššən ⬛poŋpa⬛
1sg	(N)	nə yan ⬛hea⬛
1pl	(J)	⬛nəkkən⬛ ⬛hoʔištame⬛ totoʔ
3sg	(E)	pən ⬛minpay⬛ ⬛šemhayyám⬛ yomáʔ
3sg	(I)	yomá ⬛təʔʔoyyə⬛
3sg	(R)	šiš ⬛puʔccəmmə⬛
3pl	(A)	ʔokmaŋpa ⬛haʔčukkə⬛
3pl	(D)	nə yay ⬛močukkə⬛
3pl	(T)	minnəy ⬛heyšuki⬛ šiš

There seems to be nothing in common between the verbs in each category. However, there must be some way to distinguish the subject when no explicit subject is provided, since it is explicit in the translation. We know that **šiš** in (C) means *meat*, and **nə** and **ya** from (N) have both been accounted for. That leaves only the **-šən** from (C) and **-n** from (N) as the first person marker, which is also present on the first verb in (J). It appears to be a clitic that comes immediately before the verb (the second verb when there are two) and attaches to the previous word. An analogous **-y** suffix is present in only half of the third person examples—it appears in (E), (D), and (T), but is missing from (I), (R), and (A).

To gain additional insight into this anomaly, we can examine the entire data with the knowledge of this clitic marker that precedes the verb. The sentences with first person subject are (C), (G), (H), (J), and (N). Of these, we have not analyzed (H), which has the expected **-n** suffix after **ʔu**, and (G), which instead of the clitic **-n** unexpectedly contains the *word* **də** between the two verbs. Looking for this particle elsewhere, we find it in (S), *He/she fed us*, where the object rather than the subject is first person. The same marker for both a subject and an object suggests an ergative alignment in which the subject of a transitive verb is marked ergative and the object of a transitive verb is marked the same as the agent of an intransitive verb, in the absolutive. Sentence (G) is indeed the only one that has a first person agent of a transitive verb.

This explanation is consistent with the lack of **-y** observed in sentences (I), (R), and (A), which are all intransitive—the suffix is an ergative marker. However, the second-person sentences show a different behavior. (B), (M), (P), and (Q) all contain **-m** before the verb, even though sentence (M) is intransitive. Furthermore, sentence (F), which has a second person object, is marked with **mi**. Thus, while the first and third person behave with ergative-absolutive alignment, the second person has nominative-accusative alignment.

There is one important detail remaining regarding the person markers. Only one marker is present in any given sentence, yet both subject and object markers exist for first, second, and

third person, so one must be chosen somehow and the other omitted. By organizing the transitive sentences by subject/object pairs, we can look for patterns:

1>3	2>3	3>1	3>2	3>3
(C), (H), (J), (N)	(B), (P), (Q)	(S)	(F)	(D), (E), (T)
-n (1erg)	**-m** (2nom)	**də** (1abs)	**mi** (2acc)	**-y** (3erg)

The simplest rule to describe this pattern is that Chimalapa Zoque has a "person hierarchy" in which first and second person outrank third person. A transitive sentence is marked according to the highest ranked person in the sentence.[3]

The only component of the pronouns that has not yet been explained is number (singular/plural). Whether singular or plural (or both) is marked, it is clear that the marker must go on the verb, as we have already accounted for all parts of the pronouns and no other particles are present to mark number. Looking at the verbs on the sentences with a plural first person, (G), (H), (J), and (S), we see the infix **-tam-** as a common element (except for **-dam-** in (S), the cause of which is currently unknown). **-tam-** is also in (M), the only sentence with a second person plural, so it must mark plural first or second person.

As for third person, we see **-čuk-** in (A) and (D) and **-šuk-** in (T). Although the former is more common, analyzing the latter as the underlying form explains why the verb roots **mot** (P) and **hac** (F) change to **moč** (D) and **haʔč** (A) respectively—the sequences **tš** and **cš** become **č**. This is a fairly intuitive sound change, and the pronunciation information at the bottom of the problem is helpful in seeing it (IPA: [t.ʃ], [t͡s.ʃ] → [t͡ʃ]).

This does not explain the extra glottal stop in (A), however. To account for this difference, we must note that glottal stops appear to be deleted syllable-finally, except in sentence/utterance-final position. For example, **yomáʔ** in (E) becomes **yomá** in (I), and we know that the root for *water*, ***nə**, is underlying **nəʔ** due to the appearance of **nəʔʔəm** in (P) instead of **nəm** by analogy. Thus, the verb in (F) can be analyzed as **yak + haʔc + ə**, and the glottal stop is deleted when the verb suffix is appended and the word is resyllabified.

This verb suffix, incidentally, is one that we have not yet examined. Looking at the suffixes in more detail now that we know how to separate out almost all the morphemes reveals that different suffixes are used for affirmative, negative, and dependent (e.g. *walk* in *begin to walk*) verbs:

	affirmative	negative	dependent
present	**-pa**	**-ə** (F), **-a** (N)	**-Cə**
past	**-Cə**	**-Cə**	**-e** (J), **-i** (T)
imperative	**-əʔ** (G, Q), **-aʔ** (P)	**-Cə**	**-Cə**

Notably, two of these suffixes have two possible forms, one with a schwa and one with an a. This kind of alternation suggests a form of vowel harmony, and the schwas and a's do in fact occur in complementary distribution depending on the preceding vowel. Based on the sets of vowels each occurs after ([a u] and [e o] for schwa and a respectively), it is reasonable to hypothesize that **-a** occurs after mid vowels (e, o) and **-ə** elsewhere (since a and u don't share anything in common). Similarly, **-e** occurs after a and **-i** after u, so we can guess by analogy that **-e** becomes **-i** after high vowels (i, u) pending further evidence.

Sentence (F) also contains the **yak-** prefix, which is still unidentified. The only element of the sentence unaccounted for is *making* (in *making you go into debt*), so it appears to be a causative prefix. It is also present in (S), which has the verb *feed*—implicitly a causative, as in *cause to eat*—from which we can derive the verb for *eat*, which is necessary for the questions.

---

[3]The 1>2 and 2>1 cases are not attested in the data used for the problem. The same particle, **miš**, is used for both, so I chose to leave them out to avoid ambiguity.

Another requirement for the questions is a translation of the word *there*, which does not appear in the data. However, we do have **yədə** for *this* (P) and **yəhi** and **gadə** for *that* and *here* in some order (L). By analogy, we can expect *there* to be **gahi**, as one component of the deictics appears to encode proximity (this/that) while the other classifies them (thing/place).

Looking further at sentence (L), there is again one unanalyzed affix (**-kəm-** or **-Cəm-**) and one unexplained part of the English sentence—namely, the passive voice. The only other sentence written in the passive is also the only one with the **-Cəm-** affix, sentence (R). Now is a good time to determine the underlying form of the affixes that lengthen the preceding consonant; (M) shows that **-Cəm** is really **-ʔəm**, so we can analyze the passive affix as **-ʔəm-**.

Now that most of the language's rules have been deduced, we can start systematically scanning the sentences for any missing pieces. Starting with (A), the only anomaly with the current analysis is that the verb stem for *begin* appears to be **ʔokmaŋ** in (A) and **ʔokmaw** in (G). While there is no strong evidence yet, we have never seen **w** in syllable-final or **ŋ** in syllable-initial position, so it is possible that they are the same underlying phoneme in complementary distribution.

Moving on to (B), all parts appear to be accounted for, but closer examination reveals that the word for *witchcraft* contains the verb stem **cək**, *make / do*. It also contains an unknown affix **-ha-**, which can be found in (E) as **-hay-**. The removal of the **y** between **a** and **e** is consistent with and explains the change in verb stem in (N). From sentence (E), the function of this affix appears to be to add a verb argument that specifies for whom or what the action is taking place (*for the woman*). Thus, **cək-hay-e** is *do-for-noun*, where the **-e** must convert the verb into a noun, and "doing something for someone else" is a reasonable description of witchcraft.

Sentence (C) again appears to be very simple, but keeping in mind the **ŋ∼w** alternation, the verb stem **poŋ** *burn* can be seen in the word **hopowe** *fire* from (H), clearly a related concept. It has the same **-e** suffix, but also contains the prefix **ho-**. This prefix appears in both (J) **ho-ʔiš** *examine* and (Q) **ho-ʔuk** *drink all, to the last drop*. Since **ʔiš** and **ʔuk** mean *see* and *drink* respectively, **ho-** means something along the lines of *thoroughly, completely*; a fire is something that burns thoroughly or completely. (Knowing this prefix is necessary to translate *devour* in the questions.)

(D) through (H) contain nothing of interest, but sentence (I) has the new affix **-ʔoy-**. It cannot correspond to any explicit element of the English translation, as they have all been accounted for. Implicit in the English translation, however, is that the verb *want*, which usually has an object, lacks one here. We can conclude that **-ʔoy-** is the antipassive, which removes the object from a transitive verb, common in ergative languages (note that the verb is marked with the 3rd person absolutive rather than ergative).

Sentence (J) is fully accounted for, and (K) also contains the antipassive. However, this time **-ʔoy-** + **-pa** becomes **-ʔoba**. Generalizing this change as **y** followed by a voiceless stop becoming the corresponding voiced stop allows us to fully decompose the verb in (S), which undergoes a similar process (**-ʔoy-** + **-tam-** → **-ʔodam-**) as **yak-kəš-ʔoy-tam-ʔə**.[4]

The only anomaly in the remaining sentences is that the nominalizer, which we previously described as **-e**, is in fact **-i** in sentence (O). It appears to undergo some kind of vowel harmony process, and by comparing it to the dependent present suffix in (J) and (T), we can see that it is the same one (i after high vowels (i, u) and e elsewhere). (Further evidence for the sets of vowels involved in vowel harmony is present in questions 1 through 6; e.g. we see that the nominalizer is in fact **-i** after i in question 1.)

---

[4]Note that we analyze the verbal suffix used here and for several other TAM combinations as **-ʔə** because it displays the same gemination behavior as other glottal-stop-initial affixes. In reality, the suffix is **-wə**, but the **w** behaves exactly the same way when preceded by a consonant, and there is no way to determine this from the data.