Panini Linguistics Olympiad 2016

Round 1, 9th January 2016

Solutions

Notes:

- All solutions to the problems in the PLO 2016 have been explained below. These solutions represent one of many possible ways to approach the problem and arrive at the correct answer.
- We grade your answer booklets in two rounds.
    - First, we correct the answers you wrote in your Answer Booklet. Based on the scores, we will select the top 30 to 40 senior and 25 to 30 junior candidates for grading the explanations.
    - While grading the explanations, we only look for key words, phrases or ideas in your explanations to prove that you noticed some subtle and interesting features of the language involved.
    - You are supposed to explain the rules of the language or the system you are solving. We do not look for details of how you approached the problem, and what cues you used. However, if you have not written the rules of the language, but only written your approach to the solution, we will see if you have understood the rules. If it is apparent that you have, some partial marks will be awarded.
    - **Explanations that mention the key ideas (given in blue text in gray boxes) in any format will receive full credit.**
- Please take a minute to fill up our Feedback Form at http://plo-in.org/round-1-feedback

Problem 1: Sanskrit Months

1. The phonological (or/and spelling) similarity between *Vaiśākha~Viśākhā*, *Jyeṣṭha~Jyeṣṭha*, etc. is quite apparent. This gives us the following 7 correspondences:

Month Name	Nakṣatra
Vaiśākha	*Viśākhā*
Jyeṣṭha	*Jyeṣṭha*
Śrāvaṇa	*Śravaṇa*
Āśvina	*Aśvinī*
Kārtika	*Kṛttikā*
Mārgaśirṣa	*Mṛgaśirṣa*
Pauṣa	*Puṣya*

2. It is also clear that

Bhādrapada	*Pūrva Bhādrapadā* or *Uttara Bhādrapadā*
Phālguna	*Pūrva Phalguṇī* or *Uttara Phalguṇī*

3. Doing this initial matching, one is left with 3 months: *Caitra, Āṣāḍha and Māgha.*

4. These correspondences give us the following rule for deriving the month name from the *nakṣatra* name:
    a. **All the vowels and consonants in the *nakṣatra* and month's name remains identical except for the first and last vowel.**

    b. **The first vowel of the *nakṣatra* name is transformed as follows:**
        i. ***a, ā → ā***
        ii. ***i → ai***
        iii. ***e → e***
        iv. ***u → au***
        v. ***ṛ → ār***
    c. **The last vowel of the *nakṣatra* name is replaced by *a*.**
    d. **The terms *Pūrva* or *Uttara* are not considered in the month's name.**

5. We are told that all the missing *nakṣatra* names end with *ā*, and the first vowel is not long. Therefore,
    a. *Caitra* must have been derived from *Citrā*
    b. *Āṣāḍha* must have been derived from *Aṣāḍhā*
    c. *Māgha* must have been derived from *Maghā.*

6. **We also observe that the *naksatra's* corresponding to two consecutive months (such as *Viśākhā* and *Jyeṣṭha*, *Aśvinī* and *Kṛttikā*, *Kṛttikā* and *Mṛgaśirṣa*, and *Mṛgaśirṣa* and *Puṣya*) are separated by 1 or 2 *nakṣatra's* which is not surprising as one expects the 12 month to be nearly uniformly distributed over the 27 *nakṣatra's* (so around 2.25 *nakṣatra* per month); therefore, we expect 9 of the consecutive month-naming *nakṣatra's* to be separated by 1 *nakṣatra*, and the remaining 3 of the pairs to have 2 *nakṣatra's* in between (note that both the lists are circular).**

7. Applying the above logic, since there are already two *nakṣatra's* between *Śravaṇa* and *Pūrva Bhādrapadā*, the full moon in *Bhādrapada* has to be on *Pūrva Bhādrapadā* rather than *Uttara Bhādrapadā*, which would make the gap to an unusually high 3 *nakṣatra's*.

8. **It seems that the pairs of *nakṣatra's* named as *Pūrva* X and *Uttara* X means *former* X and *latter* X, and therefore, such pairs will always denote consecutive *nakṣatra's* (in other words, there will be no other *nakṣatra* between *Pūrva* X and *Uttara* X).**

9. The *naksatra* for *Āṣāḍha*, which is one of the missing members, must come between *Jyeṣṭha* and *Śravaṇa*. The Problem says there is no missing *nakṣatra* between *Jyeṣṭha* and *Mūla.* So the missing *nakṣatra* must be between *Mūla* and *Śravaṇa.* However, by observation number 6, there must be at least one more *nakṣatra* between *Śravaṇa* and the one that's missing. **In other words, at least two *nakṣatra's* are missing between *Mūla* and *Śravaṇa*. Since it is impossible to guess the name of a *nakṣatra* which doesn't give name to a month, the only possibility is that these two missing *nakṣatra's* are: *Pūrva Aṣāḍhā* and *Uttara Aṣāḍhā* and of the two, the former lends the name to the month *Āṣāḍha.***

10. Since we recovered two of the missing *nakṣatra's*, the other two must be *Citrā* and *Maghā* without any *Pūrva* or *Uttara* counterparts. **Maghā must be positioned after Aśleṣā, which implies that Phālguna is named after Uttara Phalguṇī and therefore, Citrā must be positioned after Hastā:**

Puṣya, Aśleṣā, Maghā Pūrva Phalguṇī, Uttara Phalguṇī, Hastā, Citrā, Svāti, Viśākhā

Assignment 3:

- ***Śatabhiṣak*** → **Not possible because we do not know how to handle names ending with a consonant.**
  But guess: *Śātabhiṣak* or *Śātabhiṣaka*

Assignment 1: Missing *nakṣatra's*

Positional Number	Name of the *nakṣatra*
10	*Maghā*
14	*Citrā*
20	*Pūrva aṣāḍhā / Pūrva āṣāḍhā*
21	*Uttara aṣāḍhā / Uttara āṣāḍhā*

Assignment 2: *nakṣatra's* coinciding with full moon for each month

Month	*nakṣatra*	Month	*nakṣatra*
Caitra	*Citrā*	***Āśvina***	*Aśvinī*
Vaiśākha	*Viśākhā*	***Kārtika***	*Kṛttikā*
Jyeṣṭha	*Jyeṣṭha*	***Mārgaśirṣa***	*Mṛgaśirṣa*
Āṣāḍha	*Pūrva aṣāḍhā / Pūrva āṣāḍhā*	***Pauṣa***	*Puṣya*
Śrāvaṇa	*Śravaṇa*	***Māgha***	*Maghā*
Bhādrapada	*Pūrva Bhādrapadā*	***Phālguna***	*Uttara Phalguṇī*

Assignment 3: Deriving month-names from *nakṣatra's*

Name of *nakṣatra*	Derived Month Name
Hastā	*Hāsta*

Punarvasu	*Paunarvasa*
Śatabhiṣak	*NP*
Revatī	*Revata*

Problem 2: Aksara Sunda Kuna

We begin by observing that there are only two cities whose names end with the same sound, *Kanpur* and *Jaipur*. On the other hand, G and D in *Sundanese* end with the same symbol. This suggests that the script is written from left to right. To confirm our hypothesis, we observe that there are two pairs of words starting with the same sound, and the same is true for the Sundanese part.

Since G & D must be either Kanpur or Jaipur, and G & F share the first character – which cannot be /j/ because there is no other city name starting with /j/, G must be *Kanpur*, and  must represent the sound *k*\, and *F* must be Kochi. The last letter of F (which is /*ch*/) is similar to the first letter of A. Therefore, A must be Chennai.

Since *k*\ and *ch*\ are the only root symbols in *Kochi*, the vowels must be marked by diacritics (markings above or below a symbol that modify it, like *maatras* in most Indian languages.) We see that  represents the vowel sound *o*\,  represents *e*\ and  represents *i*\. Both *Kanpur* and *Chennai* have the symbol  . Also, *Kanpur* has the middle syllable ending with *n*\ not followed by a vowel, so Chennai also has the same feature (“*chen-nai”*). Hence, the root  must represent *n*\ and the diacritic  is used when a consonant terminates a syllable, or isn’t followed by a vowel. *k*\ in *Kanpur* isn’t marked with any diacritic, which shows that not all vowels are not explicitly marked (like *a*\).

From  at the end of option I, we can map option I to *Daman*, which is the only word ending with that sound. We can then map option H to *Delhi*, based on the first symbol.

We also notice an anomaly in the diacritic rule. The ‘r’ in *Kanpur* isn’t followed by a vowel, but the last symbol has two diacritics. This suggests that the sound \r\ is marked differently at the end of a syllable.
We see that option B has two consecutive occurrences of  , and the only city with the same vowel (since this is a diacritic, it is probably a vowel) is *Mysuru*. From this we see that  represents *u*\, so  must represent a terminal *r*\. From the new correspondences, we can identify the root symbols for *m*\, *s*\, *p*\ and *r*\ to be  , ,  and respectively.

Based on these identifications and vowel diacritics, we see that option C corresponds to *Shimla*. The lack of the symbol suggests that like *r*\, *l*\ also has a separate diacritic:  .

The character  in *Mysuru* suggests that it corresponds to *y*\, but we also note its occurrence in *Chennai*, which has the *n*\ before it marked with a diacritic. This suggests that the last syllable in *Chennai* is split into *“a-i”* and  corresponds to the root vowel *i*\.

From the identification of symbols and rules above, we can determine all correspondences.

Assignment 2:

a) \i\ appears as a root vowel in this word, and from the terminal \n\ and \d\ marked with the diacritics for \o\ and terminal \r\, we can identify this as *Indore*.

b) From the consonant root symbols for \s\, \r\ and \t\, and the diacritic for \u\, we can identify the city as *Surat*.

c) From the consonant roots for \k\, \l\ and \t\, and vowel diacritic for \o\, we can identify this as *Kolkata*.

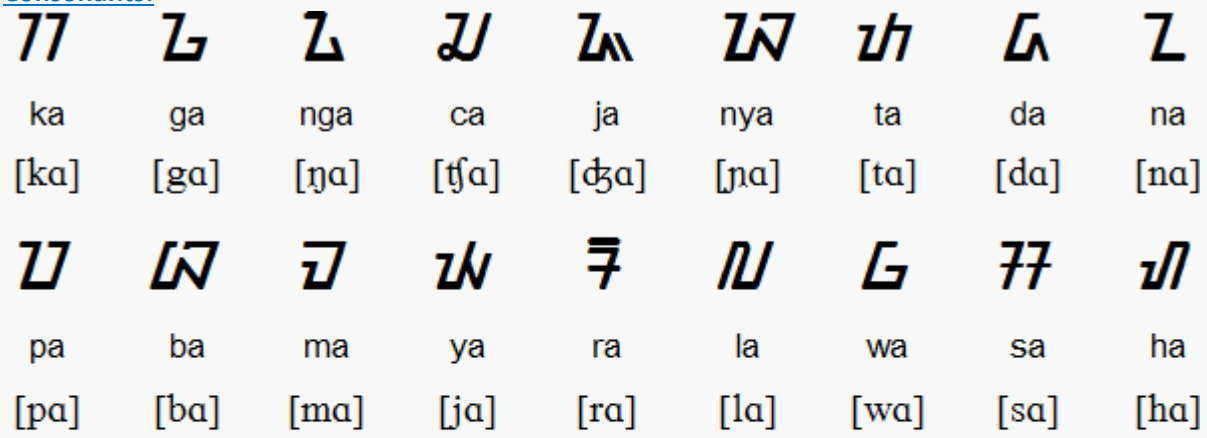
d) From the symbols for \r\, \i\ and \pur\, we can identify this city as *Raipur*.

e) Here, we encounter an unknown symbol, and hence we cannot translate it accurately. **(Full credit is given for writing *'NP'* and explaining the unknown symbol.)**
It is also reasonable to guess that the unknown symbol is */b/* and the city is *Mumbai*. **(Full credit is given for writing *Mumbai* as well)**

To summarize:

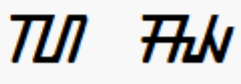
- **The script is written from Left to Right.**
- **The script consists of symbols for consonants and for free vowels. The consonants each have an inherent vowel which can be changed to another vowel or muted by means of diacritics or other modifications.**
- **Explain at least a few from the following lists.**

Consonants:

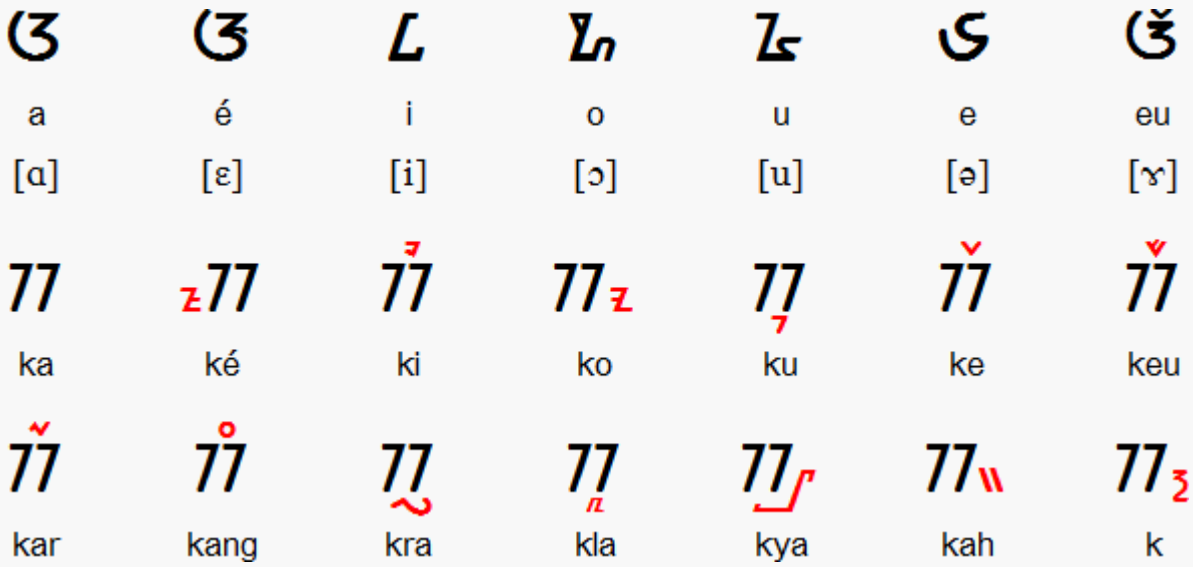


Consonants for foreign words

Additional consonants



Vowels and Diacritics:

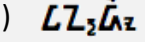
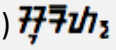
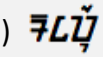
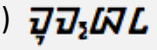


- **Assignment 2 (e):**
  - **If you answered "NP", you should have explained that the character cannot be ascertained from the data.**

Assignment 1: Correspondence

Sundanese	English	Sundanese	English
A	4	F	5
B	1	G	9
C	8	H	7
D	2	I	6
E	10	J	3

Assignment 2: Translate to English

Sundanese	English
a) 	*Indore*
b) 	*Surat*
c) 	*Kolkata*
d) 	*Raipur*
e) 	*NP / Mumbai*

Problem 3: Luganda

We observe that the Luganda words can be split into two parts (a prefix and a suffix) in such a way that we are able to isolate recurring prefixes and suffixes.

The following table classifies the Luganda words from Task 1 according to prefix and suffix:

Suffix → **Prefix ↓**	**-ntu**	**-bwa**	**-bogo**	**-wuzi**	**-gga**	**-ti**	**-ngiti**
bula-				*bula-wuzi*			*bula-ngiti*
n-		*n-bwa*	*n-bogo*				
ka -	*ka-ntu*	*ka-bwa*		*ka-wuzi*	*ka-gga*		
mu-				*mu-wuzi* *= rope*	*mu-gga*	*mu-ti*	

The following table classifies the Luganda words from Task 2 according to prefix and suffix:

Suffix → **Prefix ↓**	**-koko**	**-pisi**	**-lifoomu**	**-zzi**	**-ta**
n-	*n-koko*	*n-pisi*			
ka -	*ka-koko*		*ka-lifoomu*		
ma-				*ma-zzi*	*ma-ta = milk*

Only one Luganda word from Task 1 (viz. ***bayibuli***) can't be split unambiguously into prefix and suffix. However, we notice that ***bayibuli*** is phonetically similar to (i.e. it sounds similar to) ***the Bible***. Since the Bible is an object that has been introduced to the Luganda speakers relatively recently, it seems plausible that they wouldn't have their own word for it (just like Hindi would not have its own unique word for, say, ununpentium). By the same logic, we can guess that ***kalifoomu*** in Task 2 might be ***chloroform***.

We observe that some English words can be grouped into pairs based on a semantic (related to meaning) theme: size (big-small). For example:

1. ***dog*** - ***puppy***
2. ***river*** - ***stream***
3. ***rope*** - ***thread*** (the hint gives us: *rope* = *muwuzi*)
4. ? - ***dwarf***
5. ***hen*** - ***chick*** (from Task 2)

We observe that there is only one prefix (viz. ***ka-***) that has occurred four times in Task 1, and two times in Task 2. (However, we can rule out ***kalifoomu*** from Task 2, because of the above discussion). There are

four diminutives (AKA small objects) in Task 1, and one diminutive in Task 2. Thus, we hypothesize that the prefix ***ka-*** is used for **diminutives**.

Using this, we get the following information:

1. We are given that ***muwuzi*** = ***rope***. Therefore, ***kawuzi*** would be ***thread*** (literally "small rope")
2. The only diminutive in Task 1 without a "big" counterpart is ***dwarf***. Also, the only Luganda word with the prefix ***ka-*** in Task 1 that doesn't have a corresponding word (with a similar suffix) is ***kantu***. Thus, ***dwarf*** = ***kantu***.
3. We have two pairs of animals: ***dog*** - ***puppy*** from Task 1, and ***hen*** - ***chick*** from Task 2. We also have ***nbwa*** - ***kabwa*** from Task 1, and ***nkoko*** - ***kakoko*** from Task 2. Thus, we get ***puppy*** = ***kabwa***, ***chick*** = ***kakoko***, ***dog*** = ***nbwa***, and ***hen = nkoko***.
4. We observe that the words for the animals ***dog*** (***nbwa***) and ***hen*** (***nkoko***) both begin with the prefix ***n-***. We therefore hypothesize the prefix ***n-*** is used for **animals**. This hypothesis is confirmed by observing that ***npisi*** = ***hyena*** in Task 2, and ***nbogo*** = ***buffalo*** in Task 1.
5. This leaves us with only one remaining "big-small" pair from Task 1: English (***river*** - ***stream***); Luganda (***mugga*** - ***kagga***). Thus we get: ***river*** = ***mugga***, and ***stream*** = ***kagga***.

This leaves us with two words in Task 2: English (***water*** and ***urine***); Luganda (***mazzi*** and ***mazzi***). This means that ***mazzi*** means both ***water*** and ***urine*** (two very similar objects). The meaning would change on the context. We have been given that ***mata*** = ***milk***. Thus we get that the prefix ***ma-*** is used for **liquids.**

We now have four words from Task 1: English (***blanket***, ***blouse***, ***tree***, ***pole***); Luganda (***bulangiti***, ***bulawuzi***, ***muti***, ***muti***). Again, here we notice that ***blanket*** is phonetically similar to ***bulangiti***, and ***blouse*** is phonetically similar to ***bulawuzi***. Note that this is one of the harder parts of the problem because we have to get around three "distractors":

1. Given that blankets and blouses are common objects, one would expect the language to have its own words for them.
2. ***bula-*** appears to be a distinct prefix.
3. ***-wuzi*** has already been established as a distinct suffix.

We are left with English (***tree*** and ***pole***); Luganda (***muti*** and ***muti***). We apply the same logic that we used for ***mazzi*** to conclude that ***muti*** means both ***tree*** and ***pole*** in different contexts. Finally, we observe that the words for ***tree***, ***pole***, ***river***, and ***rope*** (***muti***, ***muti***, ***mugga*** and ***muwuzi*** respectively) all have the same prefix ***mu-***, and thus should have some semantic similarity. Again, this is one of the harder parts of the problem. They are actually all "**long objects**". Thus, the prefix ***mu-*** is used for **long objects.**

To summarize:

- **Luganda makes use of "genders" / noun classes**
- **The noun class that a word belongs to is marked with a prefix.**
- **The four noun classes are:**
  - ***n-* → animal**
  - ***ka-* → small objects (diminutive)**
  - ***mu-* → long objects**
  - ***ma-* → liquids**
- **Loanwords are present. (Some words are borrowed from English, but in phonetically modified form.)**
- **Loanwords do not take the prefix of noun classes**
- **The words *'tree'* and *'pole'* are both represented by *'muti'* possibly because they are similar in utility for the native speakers. Similarly, *'water'* and *'urine'* are both represented by *'mazzi'* and are disambiguated based on context. [*Any plausible explanation / disambiguation is acceptable.]***

Assignment 1: Correspondence

Luganda	English	Luganda	English
kantu	*dwarf*	***nbwa***	*dog*
muti	*tree/pole*	***bulangiti***	*blanket*
kagga	*stream*	***mugga***	*river*
nbogo	*buffalo*	***bayibuli***	*The Bible*
bulawuzi	*blouse*	***kabwa***	*puppy*
kawuzi	*thread*	***muti***	*tree/pole*

Assignment 2: Correspondence

Luganda	English	Luganda	English
mazzi	*urine/water*	***mazzi***	*urine/water*
nkoko	*hen*	***kakoko***	*chick*
npisi	*hyena*	***kalifoomu***	*chloroform*

Problem 4: Divehi Numerals

Problems about numbers are usually tricky because not all languages treat numbers in the same way.

For instance, English uses a *base-10* number system. This means that English has words for the 10 digits (*'one', 'two', …*), and words for some of the multiples of 10 (*'ten', 'twenty', 'thirty', …*). Combinations of these words result in different numbers. *'Twenty three'* really just means *20 + 3 = 23*.

However, languages can theoretically choose any number as a base. Many languages use *base-4*, *base-8*, *base-12*, *base 16* and even *base-60*. In this problem, you have to keep in mind that the Traditional Divehi numeral system may not follow *base-10*.

First, let's rewrite the numbers in ascending order. Notice that most of the numbers are relatively small.

3, 5, 7, 16, 17, 19, 22, 24, 34, 41, 48, 88

Observe that ***dholhas*** appears to repeat often in most of the numerals, but is not listed among the numbers. It is very likely that ***dholhas*** is a number combined very often to represent other numbers. This means that ***dholhas*** is most probably the base of this language. [Remember that ***dholhas*** itself is not present in the list, which means that it is not represented by any of the numbers above.]

Compound words could be a combination of the words for different numbers. Since ***dholhas*** usually appears as the second-to-last word ("penultimate" word), it can be assumed that the last word of any of the numbers is a digit (a number less than the base).

Here is a list of possible multiples of ***dholhas*** (appearing before the last word):

dholhas, fassihi, thin dholhas, hatheh dholhas

Here is a list of possible digits gathered from the data:

dhihaeh, faheh, hatheh, hathareh

Let's also look at how many of the Divehi numerals have just a single word. Numerals represented by just one word are either digits or multiples of the base, as described above.

faheh, fanas, hatheh, thineh, fassihi

[Notice that ***thineh*** appears to be used in ***thin dholhas*** from above. Perhaps ***thin*** and ***thineh*** refer to the same digit.]

From the above three lists, we can hypothesize that ***faheh***, ***thineh*** and ***hatheh*** represent the numbers *3, 5* and *7* (not necessarily in order), and that ***fanas*** and ***fassihi*** are multiples of ***dholhas***.

It is also a good strategy to look for patterns in the data. It is possible to assume that the remaining two of the five-single words listed above are multiples of the base. Remember that ***dholhas*** is most probably the base, and that it is not listed independently in the data.

Here are groups of multiples listed in the data:

Multiples of 3 → 3*, 24, 48*
Multiples of 4 → *16, 24, 48, 88*
Multiples of 6 → *24, 48*
Multiples of 8 → *16, 24, 48*
Multiples of 12 → *24, 48*
Multiples of 16 → *16, 48*
Multiples of 24 → *24, 48*

Since we have assumed that the base is not listed in the data, we can eliminate bases *3* and *16*.

The next part is probably best solved through a combination of intuition and brute force. If you were to assume that the base is *12*, then ***dholhas*** would mean *12*. You can verify that this makes sense by counting the number of numerals listed in the data between *12* and its next multiple *24*. There are four numerals between *12* and *24*, and there are four numerals beginning with ***dholhas***. In that case, ***fassihi*** and ***fanas*** represent *24* and *48* (not necessarily in order).

Next, notice that ***faheh*** and ***hatheh*** are digits, and that they appear in ***dholhas faheh*** and ***dholhas hatheh***. Since the numbers *17* (*12 + 5*) and *19* (*12 + 7*) are both listed in the data, ***faheh and hatheh*** represent *5* and *7* (not necessarily in order), which implies that ***thineh*** must mean *3*.

Look at ***thin dholhas faheh***. Literally, this means "*three twelve [five or seven]*". In English and in most Indian languages, some multiples of the base are represented through multiplication. [E.g. "*three hundred and two*" = (3 x 100) + 2].
Using the same logic, it is possible to assume that ***thin dholhas faheh*** means "*3 x 12 + [5 or 7]*" which could be *41* or *43*. Since only *41* is listed in the data, ***faheh*** should be the digit *5*. This means that ***hatheh*** is *7*.

Now look at ***fassihi dhihaeh***, and compare it with ***dholhas dhihaeh***. Based on our assumptions so far, ***fassihi*** either means *24* or *48*. ***dhihaeh*** is a digit (number less than 12). If ***fassihi*** meant *48*, then ***fassihi dhihaeh*** would be a number greater than *48* but less than the next multiple of 12, which is *60*. However, there is no such number given in the data.
However, if ***fassihi*** meant *24*, then ***fassihi dhihaeh*** might mean some number between *24* and the next multiple of 12, which is *36*. The number *34* appears in the data. Hence, ***dhihaeh*** could mean *10*, so that ***fassihi dhihaeh*** means "*24 + 10 = 34*".
You can confirm this by predicting the value of ***dholhas dhihaeh***. This numeral would represent "*12 + 10 = 22*". Indeed, *22* appears in the data. Hence, ***dhihaeh*** very likely means *10*.

Only two numbers remain. ***fanas*** means *48* and ***dholhas hathareh*** means *16*. Hence, ***hathareh*** means *4*.

Now we have matched all the numerals. Do you notice any patterns, or any oddities? Why do some multiples of 12 have their own word (***fassihi***, ***fanas***) whereas some multiples of 12 simply have some combination of the word for 12 and some other digit (***thin dholhas***)? You may notice that even multiples of *12* have their own names, whereas odd multiples of *12* are represented by multiplying a digit to it. To multiply a digit, the ***–eh*** suffix at the end is removed and placed before ***dholhas***.

This is why the Divehi translation of *76* from Assignment 2 cannot be determined. It must be represented as "*72 + 4*" and not "(*6 x 12) + 4*)", but the word for *72* cannot be guessed from the data.

We began with assumptions, and these assumptions could not be proved wrong and did not yield a contradiction. Hence, the assumptions must be correct and this solution must be correct.

To summarize:

- **The Maldivian duodecimal numeral system is a base-12 base-24 system, with individual words for the numerals 1 through 12 and for multiples of 24.**
- **The numeral digit roots are:**
  - **3 → *thin***
  - **4 → *hathar***
  - **5 → *fah***
  - **7 → *hath***
  - **10 → dhiha**

- The number 12 is represented as:
  - 12 → *dholhas*
- Multiples of 24 in Divehi:
  - 24 → *fassihi*
  - 48 → *fanas*

- In general, a numeral is created by adding the suffix *–eh* to the numeral root.
- Odd multiples of 12 are obtained by combining the numeral root with "*dholhas*".

  [numeral-root]-dholhas → 12 x [numeral]
- In general, to represent numbers larger than 12, two numerals are written consecutively so that they are summed.

  [multiple of 12] [numeral digit] → [numeral multiple of 12] + [numeral digit]

- **Assignment 2:**
  The number 76 must be represented as (72 + 4) and not (5x12 + 4), since 72 is an even multiple of 12.

Assignment 1: Correspondence

Divehi	Number	Divehi	Number
faheh	*5*	***thin dholhas faheh***	*41*
dholhas dhihaeh	*22*	***thineh***	*3*
fassihi dhihaeh	*34*	***dholhas hatheh***	*19*
fanas	*48*	***fassihi***	*24*
dholhas faheh	*17*	***dholhas hathareh***	*16*
hatheh	*7*	***hatheh dholhas hathareh***	88

Assignment 2: Translate to Divehi

Number	Divehi
15	*dholhas thineh*
53	*fanas faheh*
67	*fah dholhas hatheh*
76	*NP*

Problem 5: Önge

The first key to solving a problem like this (morpho-syntax, i.e. how are words formed from their elements called 'morphemes', and how they combine to make up sentences in a language) is to identify patterns.

To start off, we observe that words are sticking together to form sentences. ImagineifIwroteasentenceinEnglishlikethis! But this language seems to work this way. Such problems deliberately provide multiple sentences with similar words repeating at different positions in different sentences. This is meant to help you solve the problem. For example, in sentences #2 & #6, the word *'dog'* appears as a subject. In sentence #4, dog appears in its plural form. Note that it isn't just the plural of *'dog'*. *'All the dogs'* is different from *'dogs'* because the latter may indicate *'a few dogs, not all'*.

Now all we have to do is identify a word or group of words that occurs in all three of the sentences. We are jumping to this step only because all other elements from the sentence are totally different. Nothing but *'dog'* is common to the three sentences (except maybe past-tense markers).

The word *'ueme'* has appeared in all three sentence. For now, we will assume that *'ueme'* means *'dog'*.

To confirm the assumption, we happen to have two more sentences in English having the word *'dog'*. Look at sentences #11 and #12. Each of these sentences contains something like *'ueme'*. Now we have something to go forward with.

Notice that in sentences #2, #4 and #6, where *'dog'* is the subject (the dog is doing some action), *'ueme'* occurs at the beginning of the sentence as well. In sentences #11 and #12, *'dog'* isn't the subject of the sentence, and *'ueme'* does not occur at the beginning. Hence, it could be that the subject occurs first in *Önge* sentences.

Quickly looking at the *Önge* sentences, you may notice that sentences #3, #7 and #10 have the word *'belebe'* at the end. Let's look at the English equivalents to try and determine the meaning of *'belebe'*. Here are the components of the English sentences:

> **Subjects:** *'we', 'the father', 'the mother'*
> Maybe *'belebe'* denotes people? But then what about sentence #9, which contains *'All the women'* but does not contain *'belebe'*?
>
> **Objects:** 'matchbox', 'puppy', 'fish'
> Maybe *'belebe'* refers to anything that burns, or things that are alive? Strange and unlikely combination.
>
> **Verbs:** 'gave', 'gave', 'gave'
> This is a useful observation.

It appears that verbs occur at the end of sentences.

To confirm this assumption, check all *Önge* sentences containing the English verb *'pierced'*.

- *chogegantitebe tukirikotalotaga* (#1)
- *ueme oge*ntitebe (#6)
- *etakotalota entuementitebe* (#11)

'ntitebe' could mean *'pierced'*. Also, as a verb, it occurs at the end of sentences #6 and #11. However, sentence #1 ends with something else. Notice that before the last word in sentence #1, *'ntitebe'* appears. So what is *'tukirikotalotaga'*? Let's try to see where else it appears.

'kotalota' occurs in sentences #1, #4, #9 and #11. What is common to these sentences in English that other sentences don't have? The subjects are:

- *'All the spears'* (#1)
- *'All the dogs'* (#4)
- *'All the women'* (#9)
- *'All of us'* (#11)

Perhaps *'kotalota'* means *'all of ____'*. Then in sentence #1, *'tukiri-kotalota-ga*' could have something to do with *'all the spears'*. Now, is *'-ga'* occurring in any other sentence? Yes, in sentence #12. Notice that both the corresponding English sentences contain *'with _____'*.

It appears that *Önge* sentences with the arrangement *'with the [Noun]'* (called the instrumental case) have the arrangement *'[Noun]-ga'* at the very end of the sentence. Hence, *'with-'* word complexes occur after the verb.

This rule is consistent with our first observation in sentence #1.

Using the same rule, *'uebe'* sentence #12 seems to be the verb meaning *'covered'*. This sentence is different though. The word elements aren't sticking together like in other sentences. Here *'choge-ga'* means *'with the fish'*.

Again going back to sentence #1, we see that there is an arrangement *'chogeg-antitebe'*, and in English we have *'pierced the fish'*. So the object is preceding the verb.

After analyzing a few more sentences, we can conclude that the word order in *Önge* is *Subject-Object-Verb* (S-O-V). Other forms, like 'instrumental case markers', occur after verb. (We have data for only one such form, so it would be imprudent to make this generalization).

Now we notice that there are multiple related terms in the data, most of them human related. And not many distinct words occur so that each of them can be mapped to a single word from English.

Perhaps there are some additional prefixes/suffixes that alter the meaning?

Take two close words '*dog'* and '*puppy'*. We know from previous analysis that they both contain the word '*ueme'* for '*dog'*. Then what makes '*puppy'* different?

entueme → ent-ueme.

The sentence with 'all the women' has *allekotalota*. But we know what *kotalota* means so woman should be *alle*.

There is another sentence with 'girl'. It begins with *ent-alle*. It is consistent with our observation with '*puppy'* and '*dog'*?

Let's try to identify several more such words.

- *Father* → *onggege*
- *Old man* → *onggenge*
- *Husband* → *oge*

'*Father'* seems to be *ongg-oge*. But then why did the 'o' become 'e'? Maybe it is *ongge-oge*, not just *ongg-oge* then. We have lots of such examples in object-verb combination where the vowel is getting replaced by the previous vowel.

So *ongge-* should be the marker for *'senior'* or *'old'*. Hence *"old man"* = *ongge-onge*. Now we know what *'man'* is: *'onge'*! Can you guess what boy will be now? (Hint: small man)

It is a good observation that the sentences with the verb *'gave'* aren't agglutinating (joining together). Also, the "giver" is mentioned before the "receiver". For example,

- The father gave me the puppy (English order) → The father me puppy gave (*Önge* order)
- The mother gave fish to the father (English order) → The mother the father fish gave (*Önge* order)

So we know where to place the different objects.

To summarize:

- **Word order is SOV (Subject-Object-Verb). The secondary object occurs after the primary object, and so on.**
- **The other cases, except for Nominative and Accusative cases, occur after the verb. For instance, in *'tukirikotalotaga'*, *-ga* is the instrumental case marker.**
- **Önge is agglutinated. Complex words are formed by stringing together words and morphemes without changing them in spelling or phonetics.**
- **Vowel preference rule:**
  vowel *'e'* plus any other letter (except diphthong): *'e'* is retained, next vowel is deleted. After this, in absence of 'e' the previous vowel gets retained.
- **Explanation for Assignment 2:**
  - **c) 'NP' because the word for "*unmarried woman*" cannot be deduced without knowing word for "*wife*".**
- **(Not asked in assignment but needed to form Assignment 2 answers)**
  - **Husband - *oge***
  - **Wife - *abe***
  - **Old - *ongge***
  - **Young - *ent***
  - **Old man - *onggenge***
  - **Dog - *ueme***
  - **Fish - *choge***
  - **Matchbox - *angibete***

- **Verbs:**
  - ***ntitebe* → pierced**
  - ***belebe* → gave**
  - ***labukebe* → scolded**
  - ***uebe* → covered**
  - ***gatekkebe* → barked**

Assignment 1: Translate sentences to Önge

English	Önge
a) All the old women covered the matchbox with the spear.	*onggellekotalota angibete uebe tukiriga*
b) The big dog barked.	*ongguemegatekebe*
c) The unmarried woman gave you the fish.	***NP***
d) We pierced using all the spears.	*etia antitebe tukirikotalotaga*
e) You all drowned.	*ngikotalotatabetebe*

Assignment 2: Translate words to Önge

English	Önge	English	Önge
Old Woman	*onggalle*	Bachelor Man	*entoge*
Man	*onge*	Boy	*entonge*
Woman	*alle*	Mother	*onggebe*
Girl	*entalle*	Father	*onggege*

Problem 6: Epic Greek Poetry

Throughout this solution, lines are referenced by a letter ('C' for Correct, or 'I' for Incorrect) followed by the line number. Hence, the 3rd correct line in the data is referred to as C.3, and the 10th incorrect line in the data is referred to as I.10.

- Different people usually count syllables differently. However, it can very quickly be ascertained that each correct line of data contains almost exactly 17 syllables.
  Since this is a very consistent observation, it is likely that each line can actually be divided into 17 syllables if certain rules are followed during syllabification.

- Given below are lines C.1 and C.6 with suggested syllabification shown. (Other syllabifications are acceptable as long as the vowels in the syllable match the vowels shown here.)
  ân|dra | moi | ên|ne|pe, | Moû|sa, | po | lû|tro|pon, | hôs | ma|la | pô|lla
  âll' | ma|loud' | hôs | he|ta | roûs | ka|per | rû|sa|to, | î|e|men | ôs | per;
  A pipe ('|') is used to demarcate syllables.
  - Each syllable contains at least one vowel and might contain consonants on either side of the vowel.
  - If multiple vowels are present continuously in a single word, they are grouped together in a single syllable except when the first of the two vowels is marked by a circumflex (the 'cap').
    - According to this rule, '***moï***' and '***Moû***' are treated as single syllables.
    - In '***î|e|men***', the first two vowels are not grouped together since the first vowel has a 'cap'.

- Each metrical 'foot' contains a definite pattern of syllables. According to the problem, the first five 'feet' are identical while the sixth foot is different.
  How would 17 syllables be distributed across 6 'feet'? One possibility is for the first 5 'feet' to contain 3 syllables each, and the last 'foot' to contain a 2 syllables.

- Assume that the syllables are divided as explained in the previous point. Now look at two lines of correct data, say C.1 and C.6.
  ândra moi / ênnepe, / Moûsa, po / lûtropon, / hôs mala / pôlla
  âll' maloud' / hôs heta / roûs kaper / rûsato, / îemen / ôs per;
  A slash ('/') marks the division between two metrical 'feet'. Each of the first five 'feet' contains 3 syllables, and the last 'foot' contains 2 syllables.
  Notice how the first vowel of each foot is marked with a circumflex (a 'cap'). It can be determined that all the correct lines in the data follow this pattern.

- Hence, there are two kinds of syllables:
  - Those that contain a vowel with a circumflex ('cap'). We can call these 'heavy' syllables.
  - Those that do not contain any marked vowels. We can call these 'light' syllables.

- If 'L' represents a "light syllable" and 'H' represents a "heavy syllable", then the pattern of syllables in each line is:

  HLL | HLL | HLL | HLL | HLL | HL

  In other words, each line contains 5 feet of the form "HLL" and 1 foot of the form "HL".

- The lines I.1 through I.10 are all incorrect for the following reasons:
  - I. 1. This line is too short. The correct lines have 17 syllables each, while this has just 9 syllables
  - I. 2. This has only one vowel with a circumflex ('cap'). All the other correct lines have vowels with circumflexes repeating at regular intervals.
  - I. 3. Similar to I.2, this line has no vowel with a circumflex.
  - I. 4. This line appears to follow the *Dactylic Hexameter* as described above. However, if you divide it into feet:
    Hêliyos/ d'anôro/ ûse, li/ pôn peri/ kallêla/ limnên,
    The second foot and the fifth foot do not start with a 'heavy' syllable.
  - I. 5. Besides the number of syllables, this line does not follow the *Dactylic Hexameter* at all. It contains too many 'heavy' syllables.
  - I. 6. Like I.1, this line is too short.
  - I. 7. The first five feet of this line follows the pattern "LLH". This is not the *Dactylic Hexameter.*
  - I. 8. This line does not contain a clear pattern of 'heavy' syllables.
  - I. 9. This line is too long, and contains too many syllables.
  - I. 10. The first syllable of each word is heavy. However, this has nothing to do with the *Dactylic Hexameter*.

To summarize:

- **In the given verses, each line contains 17 syllables.**
- **The Dactylic Hexameter differentiates between two different kinds of syllables. They can be referred to as *"heavy syllables"* and *"light syllables"*.**
- **Each syllable contains at least one vowel and may contain consonants on either side of the vowel.**

- **If multiple vowels are present continuously in a single word, they are grouped together in a single syllable except when the first of the two vowels is marked by a circumflex.**
- **A syllable is 'heavy' if one of the vowels in the syllable is marked with a circumflex (like â, ê, î, ô, û). If not, the syllable is 'light'.**
- **If 'L' represents a "light syllable" and 'H' represents a "heavy syllable", then the pattern of syllables in each line is:**

  HLL | HLL | HLL | HLL | HLL | HL

  In other words, each line contains 5 feet of the form "HLL" and 1 foot of the form "HL".

Assignment 1: Do the verses follow the *Dactylic Hexameter*?

Verse Number	Y/N
a)	*N*
b)	*Y*
c)	*Y*
d)	*N*
e)	*N*