

International Linguistics Challenge

Qualification Round

2025 Edition – Solutions

Problem Sheet with Solutions

Instructions

This document contains the problems along with basic solutions.

Problem A [5 points]

An Exciting Discovery: You stumble upon a collection of mysterious audio files from an undocumented language, Language X. Along with the files, you find a note listing the audio file number and corresponding English translations.

When linguists encounter an undocumented language, they must first record and transcribe it. In this task, we will focus on the latter. Transcription is a method of writing down speech sounds in a systematic way. One way of doing this is by using the International Phonetic Alphabet (IPA), a set of symbols which uniquely map onto each distinctive sound in the world's languages.

Your task: Listen to each of the audio files linked in 1–10 and transcribe what you hear into IPA notation. Each recording consists of a single word from Language X. Use our IPA chart webpage for your transcription.

Below are two examples to help you get started. Click on the link to play the audio files. Listen and familiarise yourself with the corresponding IPA transcription.

Examples:

i. Audio 000a – 'small'
IPA: /mika/

ii. Audio 000b – 'stone'
IPA: /t͡ʃon/

Now transcribe the audio files linked in 1–10 in IPA.

1. Audio 001 – 'to sing'
2. Audio 002 – 'river'
3. Audio 003 – 'to carry'
4. Audio 004 – 'long'
5. Audio 005 – 'tree'
6. Audio 006 – 'cold'
7. Audio 007 – 'to sleep'
8. Audio 008 – 'house'
9. Audio 009 – 'to break'
10. Audio 010 – 'bright'

Why do we need the IPA?

Most of the world's writing systems are neither consistent nor have one-to-one mapping between sounds and symbols. Consider the English letter *c*. How is it pronounced in the word *concise*? The first *c* is pronounced like the *c* in *cat* or the *k* in *king*. This sound is represented as /k/ in IPA. The second *c* is pronounced like the *s* in *sea*, represented as /s/ in IPA. The English letter *c* is neither consistent in its sound nor do the sounds /k/ or /s/ uniquely map onto the letter *c*.

Moreover, writing systems are different around the world, and a single symbol can correspond to different sounds in different languages. While the letter *c* can correspond to sounds like /k/ and /s/ in English, in Turkish, the letter *c* sounds like the letter *j* in the English word *jump*.

The IPA is consistent and universal, with one-to-one mapping between sounds and symbols.

SOLUTION

1. /ʃinu/	2. /nara/	3. /plut͡ʃon/	4. /fɪnʃ/	5. /tile/
6. /gelum/	7. /burak/	8. /dreso/	9. /bradɪk/	10. /lasao/

Problem B [5 points]

The Plot Thickens: You've made progress! You now discover audio files containing full sentences in Language X. Can you decode them using what you learned in Problem A?

Identifying word boundaries is a key task in parsing language and one of the fundamental skills language learners, including children, must master. Think of your first lessons as a new language learner. The stream of speech probably sounded like gibberish, with no clear pauses between words or even sentences. As your knowledge of the language increased, you would have found it easier and easier to identify word boundaries, even if you did not know the meaning of each word. How? One of the ways language learners spot word boundaries is by using known words to identify unknown words. This is the strategy you will need to use in this task.

Your task: Listen to each of the audio files linked in 11–13. Each recording consists of a short sentence in Language X. **Transcribe the full sentence and show word boundaries clearly.** Use our IPA chart webpage for your transcription and the vocabulary from Problem A to identify word boundaries.

Here is an example. Click on the link to play the audio file. Pay attention to how a listener may identify word boundaries by using the vocabulary from Problem

A.

Example:

iii. Audio 000c

mika t͡ʃon repik
stone small roll
'The small stone rolls.'

Now transcribe the audio files linked in 11–13 in IPA. Transcribe the full sentence and show word boundaries clearly.

11. Audio 011

long river small cold stone carry
'The long river carries the small cold stone.'

12. Audio 012

small bright bird sings
'The small bright bird sings.'

13. Audio 013

tall child tree easily breaks
'The tall child breaks the tree easily.'

How do children find word boundaries?

Children are able to identify word boundaries in their native language as early as the age of 1 year. They combine probability with their knowledge of different parts of the language to do this. For instance, children learn *which* sounds can occur together in their language, as well as *how frequently* they occur together. When sounds which cannot occur together or which occur together rarely do appear next to each other, they can indicate a word boundary.

SOLUTION

Known words from Problem A mark the word boundaries; the remaining sequences are the new words.

11. fɪnʃ nara mika gelum t͡ʃon plut͡ʃon.
12. mika lasao kosap ʃinu. (new word: *kosap* 'bird')
13. gop nimu tile lorai bradik. (new words: *gop* 'tall', *nimu* 'child', *lorai*

> ‘easily’)

Problem C [5 points]

Breakthrough! Your phonetic work has paid off. Now you can start analysing the syntax and morphology of Language X.

Syntax governs how words are put together to create meaning in sentences, while morphology governs how morphemes or units of meaning are put together to create words. Linguists examine linguistic data to identify patterns, which in turn allow us to define rules. In this task, we will be looking at patterns in word order and plural noun formation.

Your task: Below are 8 sentences in Language X, along with their English translations (14–21). Analyse the data, then answer the questions in (a)–(c). Note that the translations are not word-by-word translations.

14.	kanabi ʃelu	‘The women dance.’
15.	kanabi telum feru	‘The women see the boy.’
16.	lomu telum huni	‘The dog chases the boy.’
17.	telumi lomu feru	‘The boys see the dog.’
18.	kanab kuran baraa	‘The woman studies the stars.’
19.	panon ʃelu	‘The mountains dance.’
20.	lomun pano feru	‘The dogs see the mountain.’
21.	lomun o telumi pano feru	‘The dogs and the boys see the mountain.’

Questions:

(a) What is the basic word order of subjects, objects and verbs in Language X? Give a short explanation, using examples from the dataset to support your answer.

(b) How are plural nouns formed in Language X? Give a short explanation, using examples from the dataset to support your answer.

(c) Translate the sentence below into Language X.

‘The mountains study the women and the dogs.’

Word Order Patterns Across Languages

Statistically, there are 6 possible ways of ordering Subject, Object and Verb in a sentence. One might expect all 6 of these word orders to be distributed randomly and equally across the world's languages. However, over 75% of the world's languages have either SOV or SVO word order, and only 4 known languages have OSV order. These types of distributional facts tell us that syntactic rules must be attributed to more than just probability and statistics, and may be a reflection of factors like processing ease and learnability.

SOLUTION

(a) The basic word order is SOV (Subject–Object–Verb), e.g. (16) *lomu telum huni*: *lomu* 'dog' (S), *telum* 'boy' (O), *huni* 'chases' (V). Objects are optional: intransitive sentences such as (14) and (19) are simply SV.

(b) Plurals are formed with one of two suffixes:
– **-i** is added to nouns ending in a consonant: *kanab* → *kanabi* (18 vs 14, 15), *telum* → *telumi* (16 vs 17).
– **-n** is added to nouns ending in a vowel: *lomu* → *lomun* (16, 17 vs 20, 21), *pano* → *panon* (20, 21 vs 19).
The data is equally consistent with a semantic rule: **-i** for human nouns and **-n** for non-human nouns.

(c) *panon kanabi o lomun baraa*
panon 'mountains' (S) – *kanabi o lomun* 'women and dogs' (O) – *baraa* 'study' (V)

Problem D [5 points]

The Mystery Deepens: As you examine more data, you notice something interesting: the verbs change their forms in a seemingly systematic way. What could be causing these variations?

Many of the world's languages show *agreement* between individual syntactic and morphological units. Two units agree if they have at least one feature in common. One type of agreement occurs between the verb and the subject and/or object within a sentence. Consider the English sentences *The woman dances* and *The women dance*. In the first sentence, the subject, *the woman*, is singular, and the verb must have the form *dance-s* – it is ungrammatical to say *The woman dance*. Similarly, in the second sentence, the subject, *the women*, is plural, and the verb must have the form *dance* – it is ungrammatical to say *The women dances*. This is an example of number agreement between the subject

and the verb.

Your task: Below are 4 more sentences in Language X, along with their English translations (22–25). Analyse these sentences and compare them to the sentences in Problem C (14–21). Then answer the questions in (a)–(b).

22.	kanab ʃelu	'The woman dances.'
23.	telumi lomun ferua	'The boys see the dogs.'
24.	kanab kura bara	'The woman studies the star.'
25.	lomu kuran hunia	'The dog chases the stars.'

Questions:

(a) Describe the verbal agreement system in Language X. Use examples from the dataset to support your answer.

(b) Translate the sentence below into Language X. Explain plural noun formation and verbal agreement morphology, if any, in the sentence.
Note: You will also need the vocabulary from Problems A and B.

'The bird sees the river and the trees.'

SOLUTION

(a) Verbs agree in number with the **object**: the suffix **-a** is added when the object is plural, and the bare verb is used when the object is singular, e.g. (17) *telumi lomu feru* vs (23) *telumi lomun ferua*, and (24) *kanab kura bara* vs (18) *kanab kuran baraa*. There is no person or gender agreement, and no subject agreement: the verb does not change with the subject's number, e.g. (14) *kanabi ʃelu* vs (22) *kanab ʃelu*, and (16) *lomu telum huni* vs (25) *lomu kuran hunia*.

(b) *kosap nara o tilen ferua*
kosap 'bird' (S) – *nara o tilen* 'river and trees' (O) – *ferua* 'sees' (V)
Plural: *tile* 'tree' ends in a vowel, so it takes **-n**: *tilen*.
Agreement: the object *nara o tilen* is plural, so the verb *feru* takes **-a**: *ferua*, even though the subject *kosap* is singular.

Problem E [5 points]

The Final Revelation: Your investigation has uncovered something fascinating: The data comes from two different regions, each speaking a distinct dialect of the language!

Dialects are variants of a single language, which can differ in their vocabulary, phonology, syntax, and so on. Dialects are mutually intelligible and differ from each other in systematic ways. Consider, for example, the different variants of English spoken around the world. While mutually intelligible, they each have their own distinct features, such as *la* in Singaporean English, *only* in Indian English and *innit* in London English.

Language X has two dialects with differences in their verbal agreement systems. The data you have analysed in Problems C and D is from Dialect I. In this task, you will explore dialectal differences in agreement by examining data from Dialect II.

Part (a)

Your task: Analyse the sentences from Dialect II in 26–29. Compare the agreement patterns of Dialect II to the agreement patterns you identified for Dialect I in Problem D.

26.	kanabi ferua	'The women see.'
27.	kanabi telum feru	'The women see the boy.'
28.	panon ferua	'The mountains see.'
29.	lomu panon ferua	'The dog sees the mountains.'

Question: Describe the agreement system of Dialect II. Identify one similarity and one difference between the agreement systems of Dialect I and II. Use examples from the dataset to support your answer.

Part (b)

Once syntacticians have identified rules, they use those rules to predict unseen data. You will now use your knowledge to make predictions about unseen data from Language X.

Both Dialect I and II allow pronominal subjects to be dropped. In other words, when the subject of a sentence is a pronoun (e.g. *he, she, they, it*), it can be omitted.

Dialect I:

30.	kanabi ʃelu	'The women dance.'
31.	ʃelu	'He/she/they dance.'

Question: Based on the agreement rules you have identified for Dialect II in

Problem E Part (a), what predictions can you make about the equivalent of (31) in Dialect II?

SOLUTION

(a) In Dialect II the verb agrees in number with the **object** if there is one, and with the **subject** if there is no object; object agreement takes precedence. E.g. (27) *kanabi telum feru*: plural subject but singular object, so no *-a*; (29) *lomu panon ferua*: plural object; (26) *kanabi ferua* and (28) *panon ferua*: no object, so the verb agrees with the plural subject.
Similarity: both dialects mark plural number agreement on the verb with **-a** and both have object agreement, e.g. Dialect I (23) *telumi lomun ferua*, Dialect II (29) *lomu panon ferua*.
Difference: without an object, Dialect II verbs agree with the subject, e.g. (26) *kanabi ferua*, while Dialect I verbs do not, e.g. (14) *kanabi ʃelu*, (19) *panon ʃelu*.

(b) Sentence (31) has no object, so in Dialect II the verb agrees with the (dropped) subject. Since the dropped pronoun can be singular or plural, there are two variants:
– *ʃelu* 'He/she dances.'
– *ʃelua* 'They dance.'