7 Syntax

7.1 Introduction

Syntax is concerned with the study of sentences and phrases. We remember that sentences can typically be considered to be formed from a noun phrase and a verb phrase (sentence = NP + VP)[1] (though not necessarily in that order). Therefore, syntax problems are nothing more than a noun morphology problem combined with a verb morphology problem.

Nevertheless, since NP and VP are related in the same structure, there can be some additional interactions between the two. For example, the noun phrase can contain morphemes indicating the roles of nouns in the sentence (subject, object, agent, patient, etc.). Since the arguments can be expressed though nouns (not only pronouns, as was the case in the previous chapter), there can also be a wider variety of distinctions in the verb phrase. For verb phrase problems, the principal distinctions were person (1, 2, 3), number (singular, dual, plural) and, sometimes, gender (masculine, feminine), but for syntax problems we can also add distinctions such as ±human (difference between nouns which refer to humans versus the rest).

7.2 Word order

Word order is one of the most important phenomena in syntax problems. While in the NP and VP problems the number of words is relatively small, in sentences and phrases the number of words can increase considerably; therefore the order in which these words are placed plays a very important role.

Generally, when talking about the word order of a language, we refer to the order of the subject, verb, and object; therefore, we can have six possible patterns: SOV, SVO, VSO, VOS, OSV, OVS. Cross-linguistically, the SOV and SVO patterns are the most common; they account for patterns in around 90% of the world's languages. The next one is VSO, which occurs in approx. 8% of languages (such as Classical Arabic, Tagalog, or Celtic languages like Irish, Welsh, Breton,

[1]The abbreviations used are: NP = noun phrase, VP = verb phrase.

and Manx). The least common patterns are VOS, OVS and OSV, the latter being encountered in fewer than 0.5% of languages.[2] If we look closely, we notice that the most common patterns (SOV, SVO, VSO) are those in which the subject is placed before the object, while the languages in which the subject is placed after the verb account for under 5% of the languages. Thus, it is statistically probable that in a problem the subject is placed before the object.

Moreover, it is important to mention the dichotomy between languages with fixed or rigid word order and those with a flexible word order.

Fixed word order can be directly assigned to one of the six patterns above, i.e., when a language has fixed word order, (almost) every sentence follows the same word order. All the other word orders are either ungrammatical or rarely used. Languages with fixed word order often have a simpler (inflectional) morphology in subjects and objects since the role of the arguments is well determined by their position in the sentence. Thus, if we consider the English examples *The cat saw the boy* and *The boy saw the cat*, we notice that both sentences have the same structure and the only difference is the word order. Nevertheless, there is no ambiguity between subject and object (even though they both appear morphologically identical – neither the subject nor the object is marked), since the fixed word order dictates that the subject is first (before the verb) and the object is placed after the verb.

Flexible (variable) word order, on the other hand, cannot be directly assigned to one of the six patterns above, i.e., when a language has flexible word order, not all sentences have the same word order. All six patterns (or most of them) are grammatical and commonly used. Nevertheless, there can be a dominant word order, i.e., one word order that is favoured. Unlike the languages with fixed word order, those with variable word order tend to have a much richer inflectional system. This is due to the fact that, since word order is flexible, each argument needs to be morphologically marked in order to avoid any ambiguities. Let us consider the following three examples from Romanian:

*Uriașul îl dă pe copil tatăl**ui**.*
*Tatăl**ui** îl dă uriașul pe copil.*
*Pe copil îl dă uriașul tatăl**ui**.*

All of these sentences have the same meaning ('The giant gives the child to the father.'), but the order of the arguments (subject, direct object, indirect object) differs in each case. Nevertheless, the meaning is clear and the role of each noun is

[2] Based on the data from https://wals.info/chapter/81.

well defined due to the morphological marking: the subject (*uriașul*) is unmarked, the direct object (*copil*) is preceded by *pe*, while the indirect object (*tatălui*) ends in the suffix *-ui*. In this instance, the suffix *-ui* is an example of the dative case, marking the indirect object.

Due to this morphological marking, the flexible word order does not create ambiguities in sentences such as:

Tatăl îl dă pe copil uriașului. = 'The father gives the child to the giant.'
Tatălui îl dă copilul pe uriaș. = 'The child gives the giant to the father.'
Pe tată îl dă copilului uriașul. = 'The giant gives the father to the child.'

Typologically speaking, the word order in a sentence is classified only based on subject, object, and verb. However, in writing the rules for linguistics problems, it is important to include all the components of the sentence. Thus, for the sentence *The child writes a letter with the pen*, it is not enough to write S V O, but rather we should write S V O Instr. (subject, verb, object, instrumental). Moreover, when describing the word order, we usually talk about the main constituents and not the internal structure of those constituents. This means we describe the relative positions of the subject, object, verb, instrumental, location, time, etc. As for the internal structure of the constituents, this covers the order of the noun and its modifiers (adjectives, possessives, etc.), which we can write as Adj. – Noun, or, more generally, Modifier – Noun. The reason why it is preferable to separate the two is that modifiers can occur together with different constituents (subject, object, instrumental, etc.), as in the sentence *The happy child writes a new letter with the black pen*

Thus, if we wanted to combine the two structures above (S V O Instr. and Adj. – Noun), we would have to write:

[Adj – S] V [Adj – O] [Adj – Instr]

So we would have to show that the adjective can be placed before the noun in each constituent (subject, object, instrumental). Moreover, the adjective is part of the noun phrase, so its position is strictly relative to the head of the phrase (the noun), independent of its role in the sentence (subject, object, etc.). Interestingly, in many languages the word order for the sentence is mirrored in the word order for the NP, assuming the verb/noun is the head of the VP/NP respectively. So if the sentence word order is, say verb-final, the NP order will also be head-final.

Problem 7.1

Nung (Alex Wade, UKLO 2016)

Here are some sentences in Nung and their English translations:

1. *Cáu ca vửhn nhahng kíhn.*
   'I was about to continue to eat it.'
2. *Cáu cháhn slờng páy mi?*
   'Do I truly want to go?'
3. *Cáu mi slày kíhn.*
   'I don't have to eat it.'
4. *Cáu ngám hẻht pehn tế.*
   'I did it like that just now.'
5. *Cáu tan đohc hảhn mửhng.*
   'I only see you.'
6. *Cáu vửhn nhahng bô sạhm tảhng hẻht hơn.*
   'I also continue to build the house alone.'
7. *Da kíhn!*
   'Don't eat it!'
8. *Da khải hơn!*
   'Don't sell the house!'
9. *Mửhn chống ca cháhn fải khải.*
   'Then she truly was about to have to sell it.'
10. *Mửhn mi cháhn đày non.*
    'She truly can't sleep.'
11. *Mửhn náhc-thày chống bô sạhm kíhn.*
    'Then she also just previously ate it.'
12. *Mửhng náhc-thày slờng tảhng páy.*
    'You wanted to go alone just previously.'

Problem 7.1a Translate into English:

13. *Cáu cháhn đày non.*
14. *Da páy non!*
15. *Mửhn bô sạhm mi slờng hẻht hơn mi?*
16. *Mửhn ngám bô sạhm páy hơn.*

Problem 7.1b Translate into Nung:

17. 'I wasn't about to eat it just previously.'
18. 'She didn't have to eat it alone like that just now.'

19. ‘The house truly can’t eat you.’
20. ‘Then were you also about to go just previously?’

Solution

Step 1. We notice the repetition of the first word, which we can easily correlate with the subject pronoun (*cáu* = ‘I’, *mưhng* = ‘you’, *mưhn* = ‘she’). This is also confirmed by sentence 5 where we notice that the object has the same form (therefore, S = O).

Moreover, we notice that sentences 7 and 8 both start with *da* and both are in the (negative) imperative mood. We therefore deduce that *da* is the negative imperative marker.

Step 2. In sentence 7, we have only one word remaining to be translated (*kíhn*). This must be the verb (with semantic content), so *kíhn* = ‘to eat’. This is also confirmed by sentences 1, 3, and 11.

In sentence 8, we have only two remaining words, so one must be the verb, while the other must represent the noun ‘house’. Comparing with sentence 9, we deduce that *khải* = ‘to sell’ and *hơn* = ‘house’. Moreover, we infer that for negative imperative sentences the word order is *Da* V O. Furthermore, since in sentence 7 ‘it’ is not translated, we infer that the 3SG pronominal object is unmarked.

Since we have already identified two verbs, we continue to focus on the verbs. From sentences 2 and 12, we deduce that *pảy* = ‘to go’.

Step 3. Based on sentence comparison, we can also identify most of the vocabulary, especially the adverbs: *chớng* = ‘then’, *náhc-thày* = ‘just previously’, *bô sạhm* = ‘also’, *cháhn* = ‘truly’, *tảhng* = ‘alone’.

Based on the same principle, we can identify some modal verbs: *ca* = ‘was about to’, *vữhn nhahng* = ‘continue to’.

Lastly, since we already know that *cháhn* = ‘truly’, we are left with *slờng* = ‘want to’.

Step 4. Based on the above words, sentence 6 is left with only one untranslated word, mainly *hẻht* = ‘to build’. On the other hand, we notice the same word in sentence 4, this time meaning ‘to do’ (we can assume it also represents a verb). Therefore, we deduce that in Nung, similar to many languages derived from Latin, ‘to build a house’ = ‘to do (make) a house’.

Comparing sentences 3 and 10, we infer that *mi* is a negative marker. Nevertheless, the same marker occurs in sentence 2. Since the only thing left undiscovered in sentence 2 is the interrogation, we deduce that *mi* is also an interrogative marker. Moreover, we notice that if *mi* marks a question, it will be placed at the very end of the sentence (it is rather common that the interrogative marker is placed at the end of the sentence). Therefore, *mi* represents two distinct morphemes: an interrogative marker (in which case it is placed last in the sentence) or a negative marker (in which case it is placed directly after the subject).

Sentences 3 and 10 have only one untranslated phrase, 'to have to'. Nevertheless, we notice that in Nung there are two distinct words: *slày* and *fải*. The simplest explanation for the variation between the two forms is that one is used in positive (affirmative) sentences, while the other is used in negative sentences. There can be many other explanations, such as one appears if the subject is 1SG, while the other if the subject is 3SG, but this explanation does not make sense linguistically.[3] Moreover, there are two possible interpretations regarding these two forms: either there are two completely different verbs (e.g., in English the negation of *must* is usually *not have to*), or there are two suppletive forms of the same verb, one being used in affirmative sentences and the other in negative ones.

Step 5. In sentence 10, we are left with two words: *đày* and *non*, which mean (not necessarily in this order) 'to sleep' and 'can'. Comparing the examples we have got so far, we notice that the modal verb is always placed before the semantic verb. Thus, we deduce that *đày* = 'can' and *non* = 'to sleep'.

Using a similar thought process in sentence 4, in which we need to identify the words meaning 'like that' and 'just now', we can assume that 'just now' will behave similarly to 'just previously' since they both have a similar form and meaning. Since 'just previously' appears before the verb (and even immediately after the subject), it is likely that *ngám* = 'just now' and *pehn tế́* = 'like that'.

The only words we have not identified yet are 'only' and 'to see'. Since we do not have enough details to determine which is which, we check task (a) to see whether either of them occurs in those examples, offering us

[3]The linguistic explanation for which the polarity distinction (affirmative vs. negative) is more likely to be the one that is responsible for the different verb form (rather than the subject of the sentence) is that, usually, suppletive forms are driven by intrinsic characteristics of the verb (polarity, TAM, etc.) and not by their interaction with the arguments.

additional information. Moreover, we also check task (b) to see whether we need these two words. Since these two words do not occur in any of the tasks, we can just ignore them and not try to assign them to their meaning. If, nevertheless, we would like to take a guess, we would base it on the fact that most of the adverbs are placed before the verb and that the verb is usually a single word. Therefore, we could assume that *håhn* = 'to see' and *tan đohc* = 'only'.

Step 6. Since we have identified all the vocabulary, we can solve task (a). For each of the sentences, we will first write the translation of each word (in the order in which they appear in Nung) and afterwards we write the English translation. Thus:

13. I – truly – can – sleep ⟹ 'I truly can sleep.'
14. negative imperative – go – sleep ⟹ 'Don't go to sleep!'
15. She – also – not – want – do – house – question ⟹ 'Does she also not want to build the house?'

    We take into account that 'to do a house' is translated as 'to build a house'. Moreover, we notice two things: the negation *mi* does not appear immediately after the subject as we assumed (but it still does appear before the verb). Moreover, another possible translation (and more adequate, grammatically) is 'Doesn't she want to build a house either?', since the meaning of the adverb 'also' is changed in English when in the negative (compare: *She also came – She didn't come either*).
16. she – just now – also – go – house ⟹ 'She also goes home just now'.

    In this case as well, the sentence could have been translated more literally as 'She also goes to the house just now' (and still be awarded full marks), but, generally, it is preferable to use the more natural translation.

Step 7. The only thing left to do is figure out the word order. We already know that the subject comes first. Considering the verb has a fixed position, we notice that after the verb there are only three possible morphemes: the object, the question marker, and the adverb 'like that'. We can hypothesise that the question marker is always last. Since we do not have any example in which the object and 'like that' coexist, we cannot determine

the order between them; therefore, we can assume that they occupy the same position. Consequently, we can consider the Nung word order:

Negative imperative: *Da* V O

Indicative: S [...] V O/'like that' Question

By "[...]" we mean all the adverbs and modal verbs that occur between subject and verb, whose order we are still to determine. Since we know the meaning of each of them and we are only interested in the relative order between them, we can just rewrite that part of the sentence (excluding the subject, the verb and everything after the verb). The negative imperative sentences can be excluded since their word order is clear and they do not contain any adverbs or modal verbs.

ca vửhn nhahng	'was about to continue'
cháhn	'truly'
mi slày	'not have to'
ngám	'just now'
tan đohc hảhn	'see only'
vửhn nhahng bô sạhm tảhng	'also continue alone'
chống ca cháhn fải	'then truly was about to have to'
mi cháhn đày	'truly can't'
náhc-thày chống bô sạhm	'then also just previously'
náhc-thày slờng tảhng	'want alone just previously'
cháhn đày	'truly can'
bô sạhm mi slờng	'also not want'
ngám bô sạhm	'also just now'

The second, fourth, and fifth examples can be excluded: the second and the fourth because they each have only one word, and are thus not helpful in figuring out the relative positions of the adverbs; the fifth example since we have already discussed it in Step 5 and we cannot segment it.

Moreover, since we already know the meaning of all the structures, we can just write their English translations in the order in which they appear in Nung. Thus, we get:

'was about to – continue to'
'not – have to'
'continue to – also – alone'
'then – was about to – truly – have to'
'not – truly – can'

‘just previously – then – also’
‘just previously – want to – alone’
‘truly – can’
‘also – not – want to’
‘just now – also’

What is left now is no more than a logic puzzle in which we need to arrange all of these words into an overall order which is consistent with each example above. For convenience, we refer to the words by their translations. A simple method is to start with the first word (‘was about to’). It cannot be the first one in the sentence since in the fourth row the word ‘then’ appears before it. Next, ‘then’ cannot be first since ‘just previously’ appears before it in row 6. We notice that ‘just previously’ always appears first, so we can consider it as occupying the first position. Moreover, we keep in mind that we assumed that ‘just previously’ and ‘just now’ behave similarly. Therefore, we can check whether ‘just now’ also appears in first position. It occurs in only one example and it is indeed in the first position, so we can deduce that the first position is occupied by the temporal adverb {‘just now’/‘just previously’}. Now we can delete these two adverbs from the list, as well as all examples which now contain a single word. We get:

‘was about to – continue to’
‘not – have to’
‘continue to – also – alone’
‘then – was about to – truly – have to’
‘not – truly – can’
‘then – also’
‘want to – alone’
‘truly – can’
‘also – not – want to’

Using a similar thought process, we notice that ‘then’ appears only in one example (among those left), so we can consider it to be the next in line.

So far, we have: {‘just previously’/‘just now’} ⟶ {‘then’}. We are left with:

‘was about to – continue to’
‘not – have to’
‘continue to – also – alone’
‘was about to – truly – have to’
‘not – truly – can’
‘want to – alone’

‘truly – can’
‘also – not – want to’

Now ‘was about to’ appears first, so it can be the next in line.

Note that we get the same result if we start from another word. For example, if we start with the negation ‘not’ (which appears in the second row), it cannot be the first one since, in the third row, ‘continue to’ appears before it, while ‘continue to’ cannot be first since ‘was about to’ appears before it. Thus, we again end up with ‘was about to’ as being next in line. Using the same process, we establish the overall order:

{‘just previously’ / ‘just now’} → {‘then’} → {‘was about to’} → {‘continue to’} → {‘also’} → {‘not’}

We are left with:

‘truly – have to’
‘truly – can’
‘want to – alone’
‘truly – can’

We notice that ‘truly’ appears in three out of the four remaining examples and after it we have ‘have to’/‘can’/‘want to’. We therefore deduce that the modal verbs follow it and the last word placed is the adverb ‘alone’. In order to get to this result, it is important to assume that ‘want to’ and ‘can’ behave similarly, both being modal verbs. Thus, the Nung order of adverbs/modal verbs is:

{‘just previously’ / ‘just now’}
→ {‘then’}
→ {‘was about to’}
→ {‘continue to’}
→ {‘also’}
→ {‘not’}
→ {‘truly’}
→ {‘want to’, ‘have to’, ‘can’}
→ {‘alone’}

One question that might arise is: why are the modal verbs at the end (‘want to’, ‘can’, ‘have to’) separated from the verbs ‘was about to’ and ‘continue to’, which we also referred to as modal verbs? In reality, the constructions ‘was about

to' and 'continue to' are aspectual verbs, rather than modal verbs. Therefore, in broad terms, the Nung order is: TIME → ASPECT → 'also' → 'not' → 'truly' → MODAL → 'alone'.

We need to observe that the adverb *tan đohc* = 'only' does not appear in the hierarchy above. This is explained by the fact that it appears in a single sentence and it is not accompanied by any other adverb, so it cannot be compared with any other word.

Now we can solve the task (b) and write the rules.

Rules: Word order:

Negative imperative: *Da* V O
Indicative: S [...] V O/'like that' (*mi* = Question)

[...] represents all the other adverbs/modals, which are written in the following order:

{*náhc-thày* = 'just previously'/*ngám* = 'just now'}
→ {*chống* = 'then'}
→ {*ca* = 'was about to'}
→ {*vửhn nhahng* = 'continue to'}
→ {*bô sạhm* = 'also'}
→ {*mi* = 'not'}
→ {*cháhn* = 'truly'}
→ {*slờng* = 'want to'/*fải* = 'have to'/*slày* = 'not have to'/*đày* = 'can'}
→ {*tảhng* = 'alone'}

Moreover, {*tan đohc* = 'only'} belongs to this category as well, but its place in the order cannot be determined.

Problem 7.1a

13. 'I truly can sleep.'
14. 'Don't go to sleep!'
15. 'Doesn't she want to build the house either?'
16. 'She also goes to the house just now.'

Problem 7.1b

17. *Cáu náhc-thày ca mi kíhn.*
18. *Mửhn ngám mi slày tảhng kíhn pehn tế.*
19. *Hơn mi cháhn đày kíhn mửhng.*
20. *Mửhng náhc-thày chống ca bô sạhm páy mi?*

7.3 Focusing

Focusing describes the syntactic process in which one part of the sentence is emphasized. In English, focusing can be done by intonation in speaking (compare: *He **hit** the dog* and *He hit **the dog*** In the first sentence, the emphasis is on the action of hitting, while in the latter the emphasis is on the object). Moreover, a common way to focus the object in English is by using a cleft sentence (e.g., *I saw a cat* vs. *It is a cat that I saw*).

In certain languages, focusing can be done by changing the word order (in most cases, this means that the focused part is moved nearer to the beginning of the sentence) or by using certain specific markers. For example, there might be definite or indefinite articles specific for the focused form or there can be specific morphemes which signal the fact that some words are focused. In Wolof,[4] for example, there are four sets of pronouns: subject (1SG = *man*), object (1SG = *ma*), verb-focus (1SG = *damay*), and object-focus (1SG = *laa*). The first two types (subject and object) are used by default; the verb-focus pronoun is used for the subject if the verb is emphasized, while the object-focus form is used instead of the usual object pronoun, if it is emphasized. Moreover, if the verb or the object are focused, they are moved to the beginning of the sentence. Below the four forms for 1SG and 2SG are given:

	S	O	V-focus	O-focus
1SG	*man*	*ma*	*damay*	*laa*
2SG	*yow*	*la*	*dangay*	*nga*

Let us consider the sentence 'I saw you$_{SG}$' (in Wolof, the corresponding verb is *gisoon*). We can have the following three cases:

1. Neutral sentence = no part of the sentence is focused. The word order is the default one, SOV.
   Man la gisoon. ('I saw you$_{SG}$.')
2. Sentence with focused verb = the V-focus pronoun is used to replace the subject and it is placed at the beginning of the sentence (it can be considered a focus marker), being immediately followed by the verb.
   Damay gisoon la. ('I *saw* you$_{SG}$.')

[4]This phenomenon was featured in a problem by Vlad A. Neacșu (RoLO 2019).

3. Sentence with focused object = object becomes the first in the sentence, the rest of the order is preserved.
   Nga man gisoon. ('I saw *you*$_{SG}$.')

7.4 Morphosyntactic alignment

This refers to the way in which three verbal arguments (subject of intransitive verb, subject of transitive verb, object) behave. For simplicity, in this chapter we will use the following notation: Subject of intransitive verb = Subject = S, Subject of transitive verb = Agent = A, Direct object = Object = O.

In order to better illustrate this concept, let us consider the following problem:

Problem 7.2

Morphosyntactic alignments (Vlad A. Neacșu, RoLO 2016)

The following 16 sentences represent the translations of four different English sentences into four different languages, *in random order*:

1. *Bayi jugumbil baŋgul gúdaŋgu buṛan.*
2. *'ua hi'o te tamari'i.*
3. *Bayi yaṛa buṛan.*
4. *Pol'eʔi hina nawta.*
5. *Cíq'ămqalnim peexne 'áyatne.*
6. *'ua hi'o te 'ūrī 'i te vahine.*
7. *Háma peexne.*
8. *Bayi gagara baŋgul yaṛaŋgu buṛan.*
9. *Met'aii pol'eʔ nawta.*
10. *'áyatnim peexne hámane.*
11. *'ua hi'o te tamari'i 'i te 'āva'e.*
12. *Pol'eʔi nawta.*
13. *'ua hi'o te vahine 'i te tamari'i.*
14. *Hámanim peexne hísemtuksne.*
15. *Bayi yaṛa baŋgul jugumbilŋgu buṛan.*
16. *Tsu'itsui met'ai nawta.*

Problem 7.2a Group the 16 sentences into four groups, based on the language they are in.

Problem 7.2b Group the 16 sentences into four groups, based on their meaning.

Problem 7.2c Here are eight more sentences:

17. *'ua hi'o 'i te vahine.*
18. *Met'ai pol'eʔi nawta.*

19. *Bayi gúda buṛan.*
20. *Cíq'ămqalnim peexne.*
21. *'ua te hi'o 'i te tamari vahine.*
22. *Bayi jugumbilŋgu baŋgul yaṛa buṛan.*
23. *Pol'e?i pol'e? nawta.*
24. *Peexne 'áyatnim háma.*

Out of these sentences, six are wrong. Which are these and why are they wrong?

Problem 7.2d Translate the two correct sentences from task (c) into the other three languages.

Solution

Step 1. The sentence grouping based on language can be easily done considering that all sentences in Language 1 (Dyirbal) contain the word *buṛan*, all sentences in Language 2 (Tahitian) start with *'ua hi'o*, all sentences in Language 3 (Nez-Perce) contain the word *peexne*, and all sentences in Language 4 (Wappo) end in *nawta*.

Step 2. Based on the previous grouping, we have the categories:

Lang. 1 (Dyirbal)

1. *Bayi jugumbil baŋgul gúdaŋgu buṛan.*
3. *Bayi yaṛa buṛan.*
8. *Bayi gagara baŋgul yaṛaŋgu buṛan.*
15. *Bayi yaṛa baŋgul jugumbilŋgu buṛan.*

Lang. 2 (Tahitian)

2. *'ua hi'o te tamari'i.*
6. *'ua hi'o te 'ūrī 'i te vahine.*
11. *'ua hi'o te tamari'i 'i te 'āva'e.*
13. *'ua hi'o te vahine 'i te tamari'i.*

Lang. 3 (Nez-Perce)

5. *Cíq'ămqalnim peexne 'áyatne.*

7. *Háma peexne.*

10. *'áyatnim peexne hámane.*

14. *Hámanim peexne hísemtuksne.*

Lang. 4 (Wappo)

4. *Pol'e?i hina nawta.*

9. *Met'aii pol'e? nawta.*

12. *Pol'e?i nawta.*

16. *Tsu'itsui met'ai nawta.*

We notice that in each language there is one sentence which is shorter than the others. We assume that these sentences are translations of each other, so 3 = 2 = 7 = 12.

We notice that one part of these sentences occurs in all other sentences of that language (*buṛan*, *'ua hi'o*, *peexne*, *nawta*). This represents the verb.

Step 3. Looking at the sentences in each language, we notice that in Dyirbal *bayi* and *buṛan* appear in every sentence. Expanding on this, the structure of Dyirbal sentences is:

Bayi X buṛan. ⟶ for short sentences

Bayi X baŋgul Y-ŋgu buṛan. ⟶ for long sentences

Doing the same for the other languages, we get:

Language	Short sentence	Long sentence
Dyirbal	*Bayi X buṛan.*	*Bayi Y baŋgul Z-ŋgu buṛan.*
Tahitian	*'ua hi'o te A.*	*'ua hi'o te B 'i te C.*
Nez-Perce	*M peexne.*	*N-nim peexne P-ne.*
Wappo	*R-i nawta.*	*S-i T nawta.*

Step 4. By analysing the nouns in sentences 2, 3, 7, 12, we know that *yaṛa* = *tamari'i* = *pol'e?i* = *háma*.

We notice that, in each language, these nouns appear in two more sentences. Therefore, each language has a sentence which does not contain these words, so 1 = 6 = 5 = 16.

Each of these sentences contains two nouns: one which occurs in one more sentence, and one which does not occur in any other place. Checking the sentences in which that noun also occurs, we deduce 15 = 13 = 10 = 9 and *jugumbil = vahine = met'ai = 'áyat.*

We are left with the other noun from sentences 1 = 6 = 5 = 16, so *gúda = 'ūrī = cíq'ămqal = tsu'itsu.*

Now we are left with only the sentences 8 = 11 = 14 = 4 and the nouns *gagara = 'āva'e = hísemtuks = hina.*

Thus, the sentences are:

Sentence A:

3. *Bayi yaṛa buṛan.* (Dyirbal)
2. *'ua hi'o te tamari'i.* (Tahitian)
12. *Pol'e?i nawta.* (Wappo)
7. *Háma peexne.* (Nez-Perce)

Sentence B:

1. *Bayi jugumbil baŋgul gúdaŋgu buṛan.* (Dyirbal)
6. *'ua hi'o te 'ūrī 'i te vahine.* (Tahitian)
16. *Tsu'itsui met'ai nawta.* (Wappo)
5. *Cíq'ămqalnim peexne 'áyatne.* (Nez-Perce)

Sentence C:

8. *Bayi gagara baŋgul yaṛaŋgu buṛan.* (Dyirbal)
11. *'ua hi'o te tamari'i 'i te 'āva'e.* (Tahitian)
4. *Pol'e?i hina nawta.* (Wappo)
14. *Hámanim peexne hísemtuksne.* (Nez-Perce)

Sentence D:

15. *Bayi yaṛa baŋgul jugumbilŋgu buṛan.* (Dyirbal)
13. *'ua hi'o te vahine 'i te tamari'i.* (Tahitian)
9. *Met'aii pol'e? nawta.* (Wappo)
10. *'áyatnim peexne hámane.* (Nez-Perce)

Step 5. In the table at Step 3, we marked each noun with different letters, not knowing in which order they occur in the long sentences. Now, based on the correspondences, we deduce the final sentence structure:

Language	Short sentence	Long sentence
Dyirbal	*Bayi S buṛan.*	*Bayi O baŋgul A-ŋgu buṛan.*
Tahitian	*'ua hi'o te S.*	*'ua hi'o te A 'i te O.*
Nez-Perce	*S peexne.*	*A-nim peexne O-ne.*
Wappo	*S-i nawta.*	*A-i O nawta.*

We can easily solve tasks (c) and (d):

Solution 7.2c

17. *'i* occurs
18. Word order
20 Ending of the first word
21 Word order
22 Word order
24 Word order and ending of last word

Solution 7.2d

	19.	23.
Dyirbal	*Bayi gúda buṛan.*	*Bayi yaṛa baŋgul yaṛaŋgu buṛan.*
Nez-Perce	*Cíq'ămqal peexne.*	*Hámanim peexne hámane.*
Tahitian	*'ua hi'o te 'ūrī.*	*'ua hi'o te tamari'i 'i te tamari'i.*
Wappo	*Tsu'itsui nawta.*	*Pol'eʔi pol'eʔ nawta.*

This problem allows us to better understand the fundamental difference between morphosyntactic alignments across languages. Returning to the general structure of the above sentences, we notice we have four situations:

1. In Wappo, *S* and *A* are marked identically (using the suffix *-i*), but differently from *O*. In this case, we talk about a NOMINATIVE-ACCUSATIVE ALIGNMENT. This is the most common type of alignment. In this case, *S* and *A* are the nominative arguments, while *O* is the accusative argument.

2. In Nez-Perce, *S*, *A* and *O* are all marked differently. This is, by definition, the TRIPARTITE ALIGNMENT.
3. In Tahitian, there is no difference between *S*, *A* and *O* (all three are unmarked), so we say that this language features a DIRECT ALIGNMENT.
4. In Dyirbal, *S* is marked like *O* (in this case, unmarked or using a null morpheme), but differently from *A* (which receives the suffix *-ŋgu*). This language exhibits an ERGATIVE-ABSOLUTIVE ALIGNMENT. The agent is the only ergative argument, while the absolutive arguments are the subject and the object.
5. There is another type of alignment, extremely rarely used: the TRANSITIVE alignment in which *A* and *O* are marked the same, while *S* is marked differently.

A schematic representation of the five types of alignments is shown below, where identically marked arguments are highlighted in the same colour:

Alignment	Nom-Acc	Erg-Abs	Tripartite	Direct	Transitive
S					
A					
O					

It is important to understand that the morphosyntactic alignment is intrinsic to the language. Thus, in English we cannot talk about an ergative argument simply because English does not follow an ergative-absolutive alignment. Thus, the first step is determining the type of alignment that the language follows.

In the solution to Problem 6.13, we mentioned that the plurality of the verb is determined by the subject of the intransitive verb or the object of the transitive verb. In this problem, the two arguments have the same role (determining the verb plurality), so we can combine them under the specific of the absolutive case, stating that the marker *-pa* appears if the absolutive argument of the verb is plural.

7.5 Split alignment

Certain languages can have two (or more) types of alignments, each of them appearing in a specific linguistic context. For example, Pashto has a split align-

ment: it follows a nominative-accusative alignment in the present tense, but an ergative-absolutive alignment in the past. We can compare the following examples:

Ze wlarrem.	Dai *me* woleed.	*Ze* yay woleedelem.
I went	him I saw	me he saw
'I went.'	'I saw him.'	'He saw me.'
Ze dzem.	*Ze* yay weenem.	Dai *me* weenee.
I go	I him see	he me sees
'I go.'	'I see him.'	'He sees me.'

We notice that 1SG can be expressed in two ways: *ze* and *me*. In the past tense sentences (first row), *ze* is used for subject and object, while *me* is used for agent. Thus, the past follows an ergative-absolutive alignment, with *me* being used as the ergative form of 1SG and *ze* the absolutive form. At the same time, in the present tense sentences, we notice that *ze* is used for S and A, while *me* is used for O. Thus, the present tense follows a nominative-accusative alignment, with *ze* being the nominative form of 1SG and *me* the accusative form.

Another situation in which we can talk about the coexistence of two different alignments in the same language (but which is not considered split alignment) is that in which, historically speaking, a language had a certain alignment but, in time, due to changes in morphophonology, it came to use the direct alignment (completely unmarked). This can be easily observed in English, where we talk about a nominative-accusative alignment of the pronouns (*he – him*, *I – me*), but about a direct alignment of the nouns.

In linguistics problems the ergative-absolutive alignment is commonly found (together with the nominative-accusative one), while the split alignment is usually highlighted between these two types of alignment. Although in Pashto the context of the two alignments depends solely on the tense, the distinction is motivated for other reasons too: person, discourse prominence of arguments (noun vs pronoun), etc. Every time we see both transitive and intransitive sentences in a linguistics problem, we need to take into account the possibility that the language has a different alignment. Although nominative-accusative languages are more common, especially in the West, and therefore more familiar, ergative languages represent about a quarter of all world languages, and are particularly found in less known language families, and so are proportionally more likely to occur in linguistics problems! Another feature of ergative languages is that they are almost all verb-initial or verb-final, almost never SVO.