

# Squib Three: Kazakh Syntax

Aidan Mokalla, Sara Tapalova

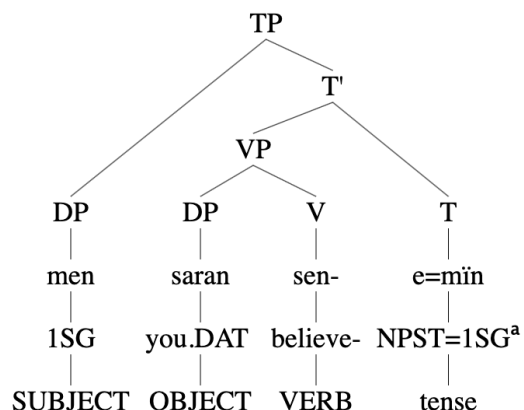
## I. Introduction

This squib examines, constructs, and motivates various claims about Kazakh syntactic structures. Its unique syntactical framework is characterized by its predominant Subject-Object-Verb (SOV) order, an attribute that serves as the foundation for our analysis. Through a detailed examination of basic word order, noun phrase (NP) and verb phrase (VP) structures, and other phrase structures, we hope to express the syntactic rules and patterns that define the Kazakh language in the simplest possible terms. We use tools like examining transformations and the implications of head-directionality, head- and dependent-marking, and other typological parameters to search for insights into its syntactic peculiarities and universalities.

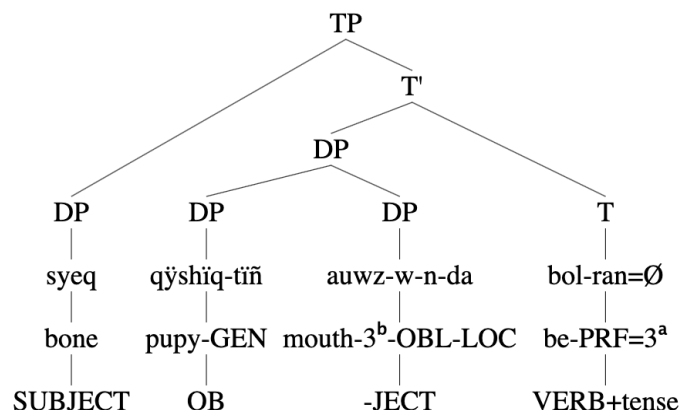
## II. Basic Word Order

The Kazakh language exhibits universal SOV word order. Kazakh tense phrases are head-final, so SOV is always followed by syntactic tense, as indicated in our trees. Consider the following:

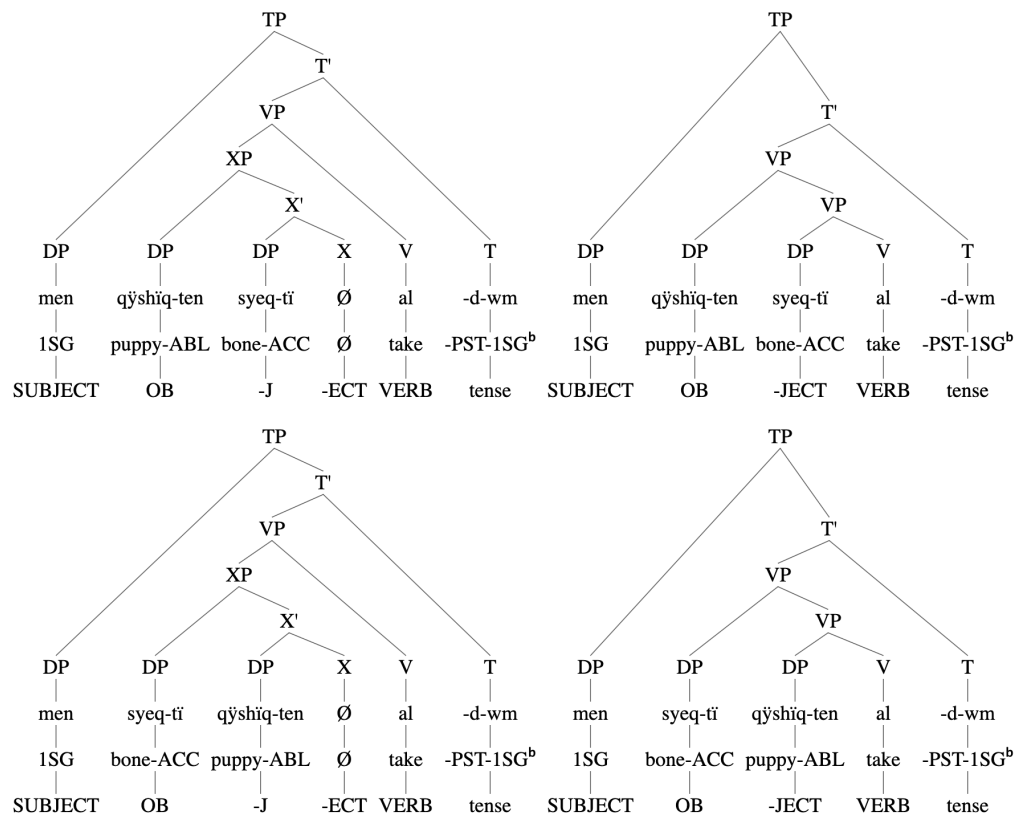
1) “I believe you” <U832>:



2) “The bone had been in the puppy’s mouth” <1589>:



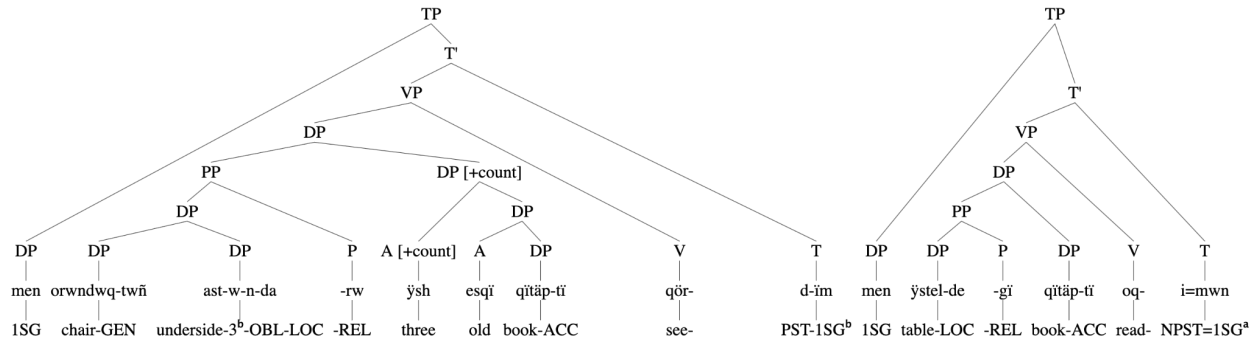
Regardless of any controversy regarding decisions about constituencies in the above trees, which will be motivated later, it is clear that Kazakh has consistent SOV order. In the following example, we have four possible trees, depending on whether we prefer X-bar theory or repeated DP selection, and depending on the order of our verbal arguments. Again, all else aside, we observe consistent SOV patternings:



Kazakh nouns (DPs) are not marked for determination. As such, for the sake of our syntactic analyses, we will consider all nouns in Kazakh to be *inherently* ‘determined.’ As a result of this assumption, there is no need to mark Kazakh nouns for determination in the first

place. DP elements can select the following: other DPs, complementizer phrases, and adjectives/adjective phrases. Postpositional phrases selected by DPs take a relativizer:

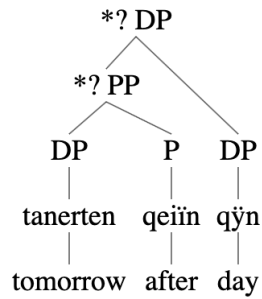
4) “I saw three old books under the chair” <U1810> & “I read the book on the table” <U1564>:



There is, of course, the question of whether the relativizing suffix <-gĩ/-rw> in the above examples really can be the head of a postpositional phrase; The same question arises, that of if the DP elements are selected by the V head sequentially or at the same point in the generation. Crucially, however, the head of the phrases <ýstelde qĩtäptĩ> and <orwndwqtwñ astwndarw> are uncontroversially at the end of the phrases, regardless of the phrases' statuses.

Unfortunately, upon review, there is not a single example in our database of a postpositional phrase headed by one of the four *lexical* postpositions we have discovered (<qeiĩn>, <turalw>, <ýshñ>, and <arqwlw>) being selected by a DP (there are examples of these phrases being selected by other types of constituents which we discuss later). However, since we have seen that these four words *can* be the head of postpositional phrases (see example (19)), we hypothesize that these phrases can be selected by following DPs. So, the following head-final construction would be a grammatical translation of the noun phrase “the day after tomorrow”:

5) ?\* “the day after tomorrow”:



The PP head-finality predicted above will be discussed in a later section. DPs selection of other DPs in Kazakh also obeys head-finality. Consider the following compound noun, where the specifier noun comes first and the head second:

6) <L56>:

‘aue rae’

aue rae

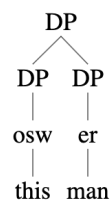
air state

“the air’s state”

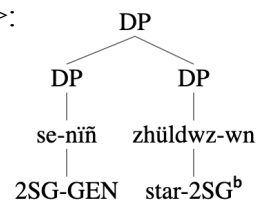
‘Weather’

Composed DPs with complex internal syntactic structure are also head-final. That is, as we would expect with other DP-DP selections, all demonstratives within DPs, like deictics and possessors, also demonstrate head finality:

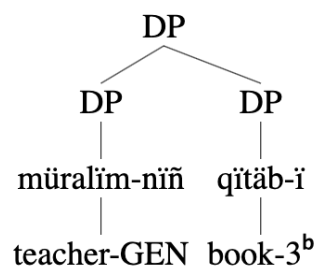
7) “this man” <525>:



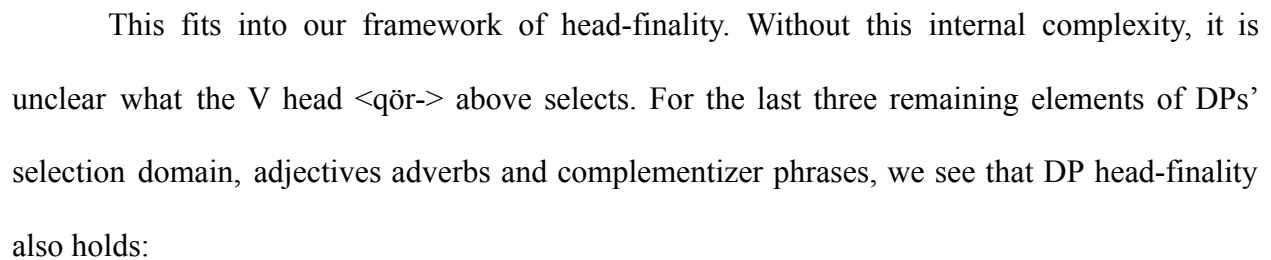
8) “your star” <513>:



9) “the teacher’s books” <1572>:



10) “I saw both you and the horse” <U1749> and “me and you” <U1756>:



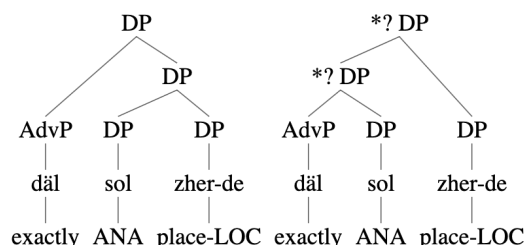
```

graph TD
    TP --> DP1[DP]
    TP --> T
    DP1 --> men-de[men-de]
    men-de --> 1SG-LOC[1SG-LOC]
    T --> DP2[DP]
    T --> DP3[DP]
    DP2 --> ysh[ysh]
    ysh --> three[three]
    DP3 --> A1[A]
    A1 --> yiqen[yiqen]
    yiqen --> big[big]
    DP3 --> DP4[DP]
    DP4 --> A2[A]
    A2 --> qarw[qarw]
    qarw --> old[old]
    DP4 --> DP5[DP]
    DP5 --> A3[A]
    A3 --> ädemi[ädemi]
    ädemi --> beautiful[beautiful]
    DP5 --> DP6[DP]
    DP6 --> DP7[DP]
    DP7 --> zhwqlw[zhwqlw]
    zhwqlw --> horse[horse]
    DP6 --> T2[T]
    T2 --> bar[bar]
    bar --> exist[exist]
  
```

[illegible]

Interestingly, in this example, the sister of the complementizer phrase, *book*, must move out from within the phrase to select it. By topicalizing in this way, the DP maintains head-finality.

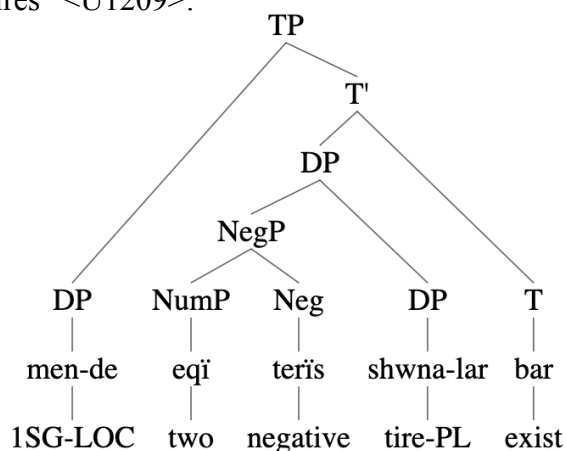
13) “at that exact place” <U1373> and \*? “at exactly that place” (adverb phrase selection):



Thus, we have demonstrated that Kazakh DPs are always head-final, across all types of elements that DPs can select.

In the authors’ last squib, we claimed that a plural “noun is marked for plurality if and only if it isn’t counted explicitly, or if it’s counted zero” (page 2). I continue to claim that the proposed broader restriction of the numeral “not co-occurring with a noun” is not a strict enough criterion:

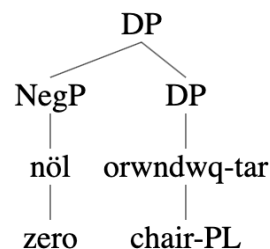
14) “I have negative two tires” <U1209>:



However, the claim from our last squib does indeed seem to be too broad (see example 7 — we would expect a plural marking to be possible). We propose instead that a plural noun is never marked for plurality, unless it directly selects a “negative phrase”, in which case it is

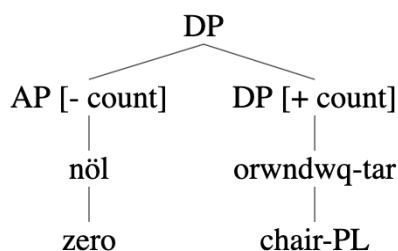
marked for plurality. We can envision the NegP as a number phrase that has been marked semantically for negation. We can state our observation as the result of a syntactic rule: That a number phrase cannot be selected by an element that is already marked for count, such as a DP with a -PL marking, and as a result, there must be an intervening element that is not marked for count, like a Neg head above, intervening between a number phrase and its predicate to make the marking possible. This would explain why we don't see a -PL marking in example (11), since the number phrase is directly selected by the DP. If we consider “zero” to also be one such Neg phrase, as it negates the existence of its predicate's count, then the following construction is what we would expect, and is what Sara generates:

15) “zero chairs” <U549>:

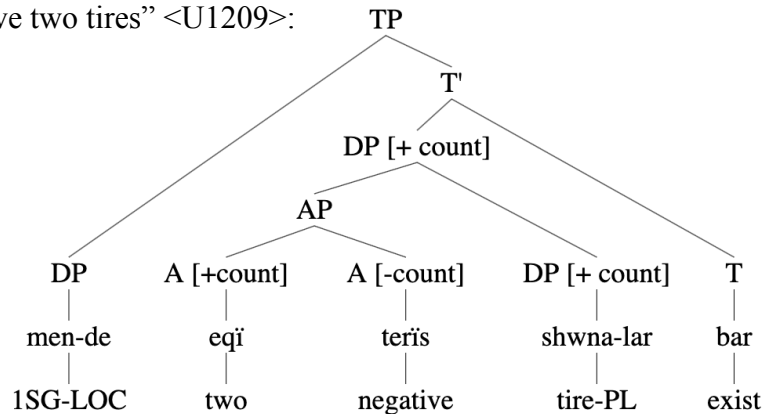


An alternative analysis of this phenomenon, if we're skeptical of the idea of a “negative count phrase,” is that adjective and noun phrases can be marked by a “count” feature, either [+count] or [-count], and that elements with the same count feature cannot merge. This analysis, like the prior, also explains why 2 sequential adjectives (e.g. “negative two”) can be a constituent in these examples, but cannot be a constituent in example (11) (e.g. “big old”): The [+count] and [-count] features attract one another, canceling out, which leads to earlier merging. This is similar to how the [+GEN] and [+agr-B] features cause DPs to merge early in example (2).

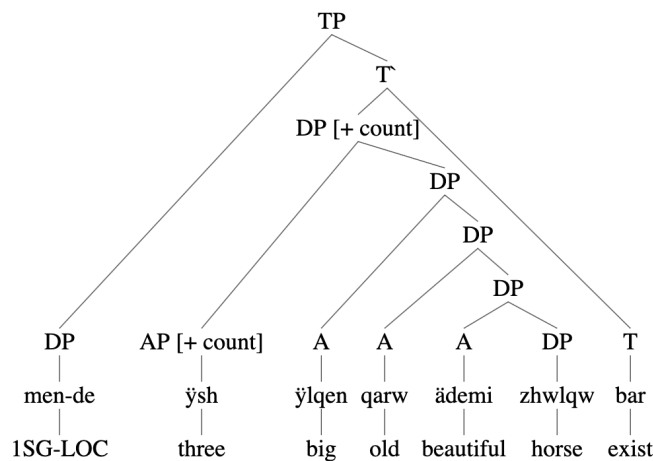
16) “zero chairs” <U549>:



17) “I have negative two tires” <U1209>:



18) “I have three big old beautiful horses” <U1282>:



#### IV. Verb Phrase Structure

Morphological verbs in Kazakh, such as <bar> in example (above), <zhazuwm> in example (12), or <senemĭn> in example (1), are related, but not the same, to the head of syntactic verb phrases. As we have seen, morphological verbs in Kazakh always carry tense information. Sometimes, morphological verbs carry only tense information, and no semantic information, as in examples (18, 14, etc). These verbs, or rather these *copula*, are addressed below, where we discuss T-constructions, and their relationship to verbs, in greater detail. For now, these two claims should not be incredibly surprising, given that English has the same system.

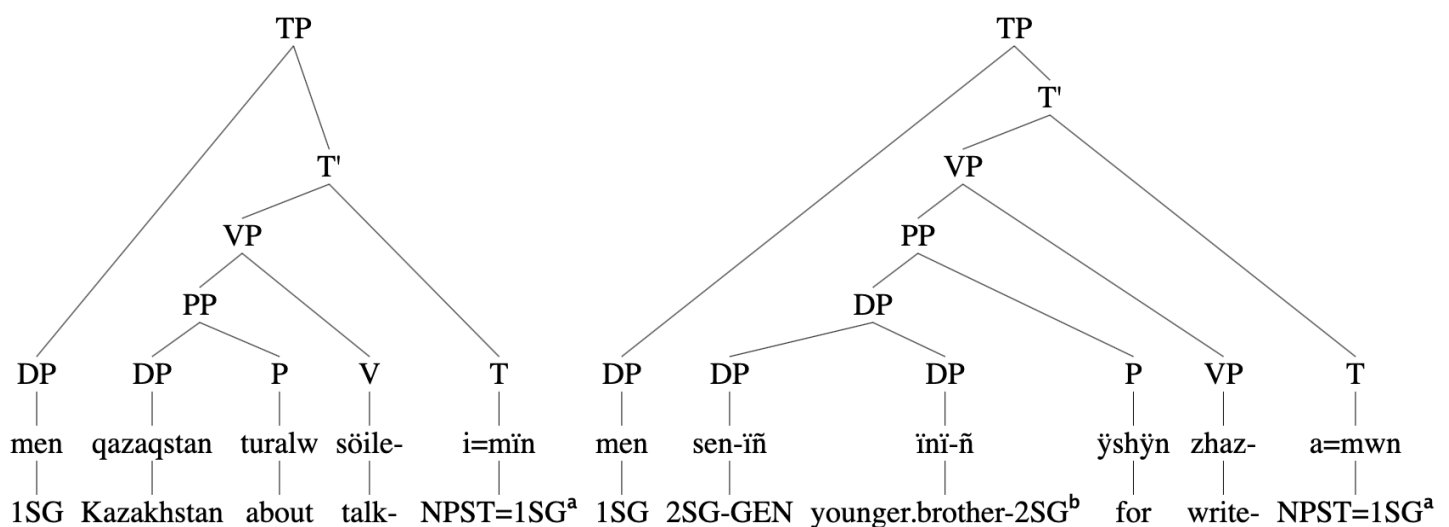


Kazakh verbs and verb phrases can select the following elements: DPs, PPs, adverbial constructions (AdvPs), other verbal elements, and perhaps X phrases (as in example 3). When selecting both direct and indirect objects, direction is only differentiated by morphological case markings, and is not indicated by word order / syntactic structure (3). This can also be explained in terms of a strong head-finality rule: The V head, which wants to select an argument, must select its preceding element, forming a VP. There is no reason this element must be the direct or indirect argument. However, if whichever argument remains unselected is of the same constituency as the first, it must merge with the following VP to form a larger verb phrase. Since both arguments can be syntactically selected first indiscriminately, VP argument direction must be determined morphologically.

When verbs select postpositional phrases, they are head-final:

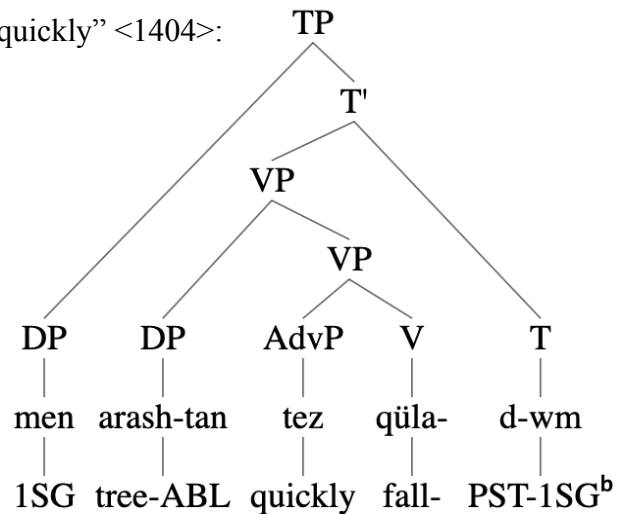
19) “I’m talking about Kazakhstan” <1138>,

“I’m writing (this) for your younger brother” <U1155>:



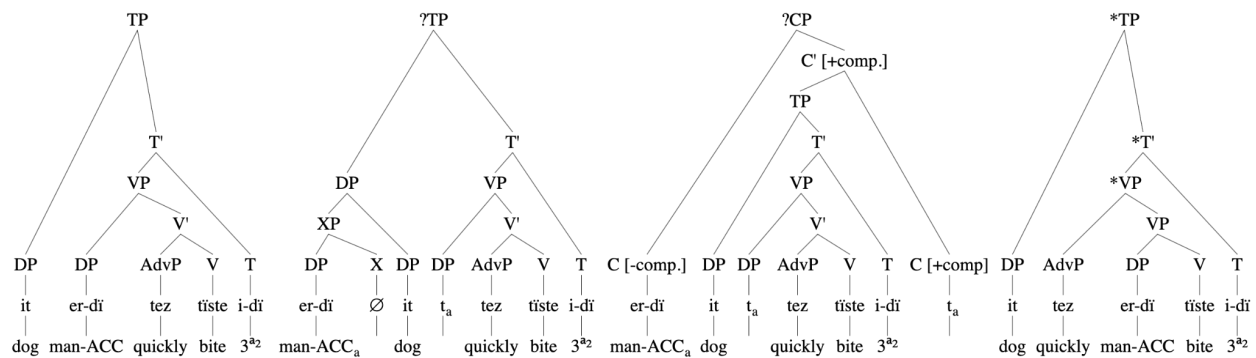
When verbs select adverbs, they are yet again head-final:

20) “I fell from the tree quickly” <1404>:



It seems that verbs can select adverbs, but entire verb phrases cannot. In the second, third and fourth trees below, we see that the syntax is more willing to generate a much denser construction than it is willing to violate this rule. We again have two options regarding X phrases as well:

21) “The dog quickly bites the man” <U1678, U1679, U1680, U1681>:



This might indicate an interesting morphosyntactic relationship where morphological case is first assigned to the dependent of the top verb phrase, then to the dependent of the second-top, etc.

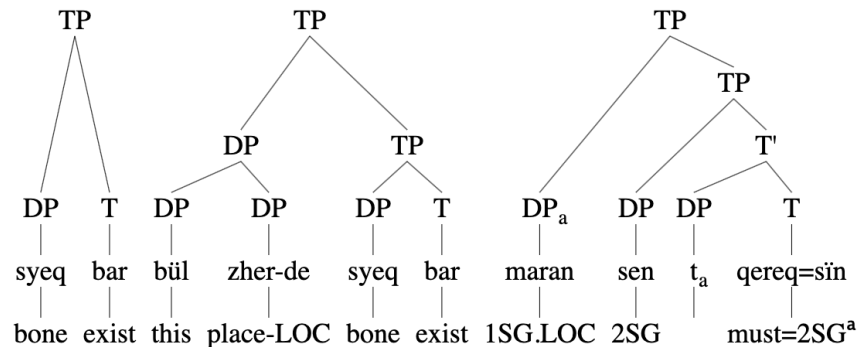
Why might this be? Consider the following:

- A. The T head must assign a subject role if the T phrase has a semantically specified subject.
- B. All subjects must begin in the T head to receive the subject role
- C. All DPs must receive case in Kazakh
- D. DPs can only be assigned case when if they are either:
  - a. A specifier of a T head (nominal case),
  - b. Being C-commanded by a T head (morphological case), or
  - c. Move through/into one of these positions.

So far, we have not encountered any obvious violations of these rules. In fact, many of the earlier rules we have discussed can be reduced simply to consequences of the above:

- A. If the subject starts in the T head, and TPs are head final, then the subject must move to before the entire T/T' constituent. This generates the subject-first rule.
- B. We have seen that subject agreement marking occurs in the T-head. This is where the subject starts before movement, but since it cannot receive case in this position, it can leave behind an agreement marker, but it has to move up until it is either a specifier or being C-commanded by a T head.
- C. Without even a verb, T can merge with DP to become TP when there is no subject role to assign ((22)(1, 2)). This is because these first two sentences satisfy the subject assignment requirement, whereas a *semantic* verb (but not <bar>) would need to interact with a subject role. Because of this, when T selects VPs, it can only partially project to a T', and never has the option to fully project to a TP because it has not yet been able to assign the subject role. In the third tree below, even though there is no subject, the order of the first two words cannot be switched because then <maran> would be marked for case twice:

22) “There is a bone”, “There’s a bone here” <U694>, “I need you” <U1808>:

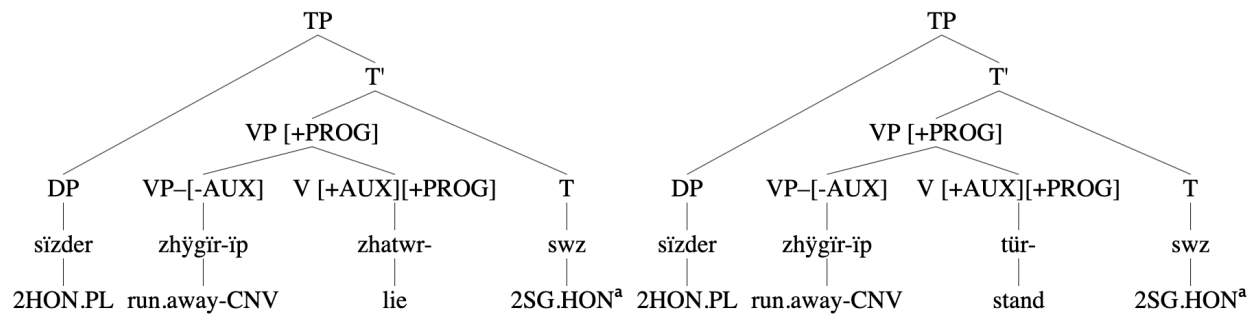


- D. This is precisely why the T head sometimes has person agreement, and why it sometimes does not: If there is a DP with a semantic subject role, then it will start out in the T head position, before moving to subject position and leaving behind an agreement marking. In the above example however, there is not a DP argument with a subject role. So, since 1SG is already assigned morphological case (which is allowed, since its position is the specifier of a T phrase), and 2SG still needs to be assigned case, 2SG starts out in the T head, before moving from head to specifier so that it can be assigned (“nominal”) case.
- E. In example (21(3)), it makes more sense that the subject start in the T head than in the final C head, because under this reading, the subject can receive the [+complementizer] feature directly from C position without having to abstract that feature.
- F. We never observe agreement marking on the copula and morphological case marking on the subject at the same time. This is because either:
- The sentence has a semantic subject role, and so the subject originated in the T head, and the T head is then marked for agreement with the subject, while the subject receives nominal case and is thus unmarked. Or,

- b. The sentence does not have a subject role, so the T head is not marked for agreement, and any DPs without case can move to TP specifier positions to be assigned case. If no DPs need case, an adjective can specify the TP, as in (25).

Finally, verbs and verb phrases may select other verbs, as we've seen with progressive/current constructions:

23) “Y'all (HON) are running away” <U581>:

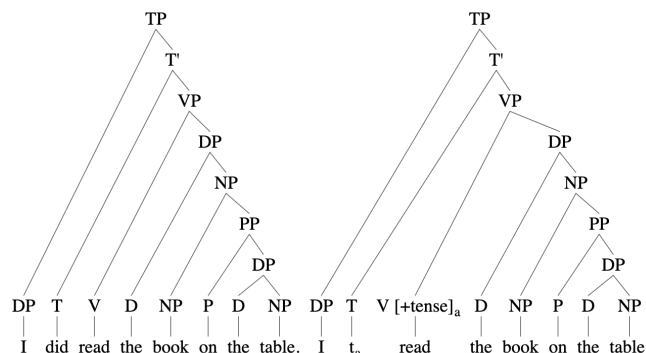


Thus, we have demonstrated that Kazakh verb phrases are always head-final, across all types of elements that verbs and verb phrases can select.

## V. Other Phrases

Earlier, we discussed some similarities between English and Kazakh tense constructions. Similarly as well, when a verb expresses both semantic information and tense, an overt, distinct copula just isn't needed, like in examples (4, 23). Consider the English tree of 2 possible translations of (4):

24) See (4)

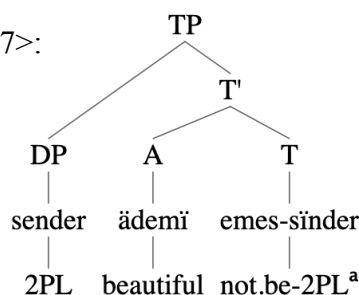


However, this alternation isn't possible in Kazakh, since V and T are always one morphological unit. This means that verbs must be marked for tense at a morphological level, unlike in English, where verbs can often be marked for tense at either a morphological or syntactic level. When the tense feature is transferred in English, the V head must be C-commanded by T to get this feature.

This stricter construction in Kazakh can be explained as the result of syntactic head finality. Because the T head always selects a VP (or DP), and because TPs are head-final (as far as we have seen), the T head must always be directly following the head of the VP that it selects. So, as we mentioned earlier on page 1, it is unsurprising that a verb root and the elements that necessarily always directly follow it be parsed as a single morphological unit. As we'd expect, when the T head has no semantic verb preceding it with which to form a morphological verb (as in example (18)), then a morphological copula is generated. No semantic verb means no subject role to assign, which means that this morphological copula verb won't start out with a DP to give an agreement value.

When tense heads like <emes> can select adjectives, and these constructions are yet again still head-final:

25) "y'all aren't beautiful" <U1717>:

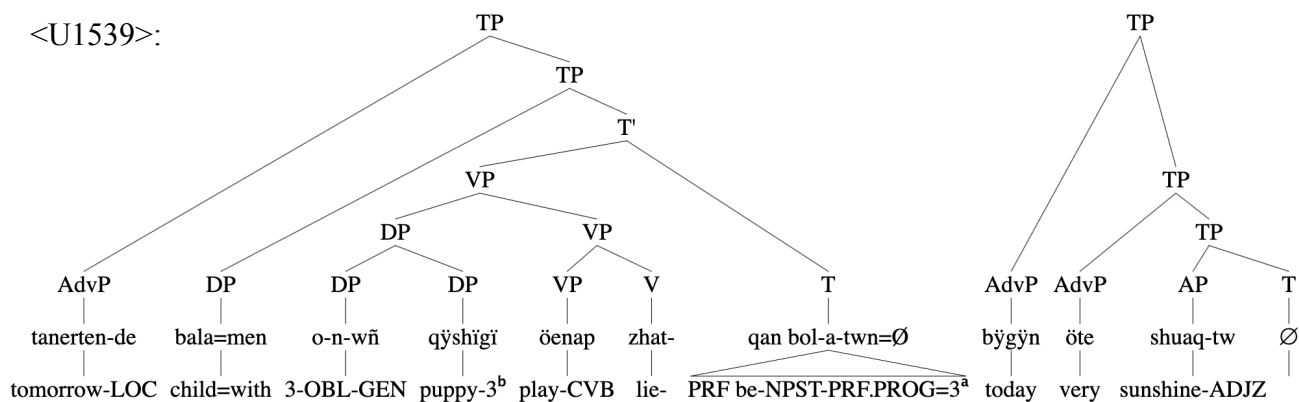


In the trees above, we have made the claim that tense phrases do not necessarily contain verb phrases at all (16, 23). This is because we claim T heads may directly select DPs, just as TP heads may. According to optimality theory, removing the stipulation that only TP heads may

select DPs, while T' heads cannot, is easily motivated by the system's goal to simplify underlying head selection criteria as much as possible.

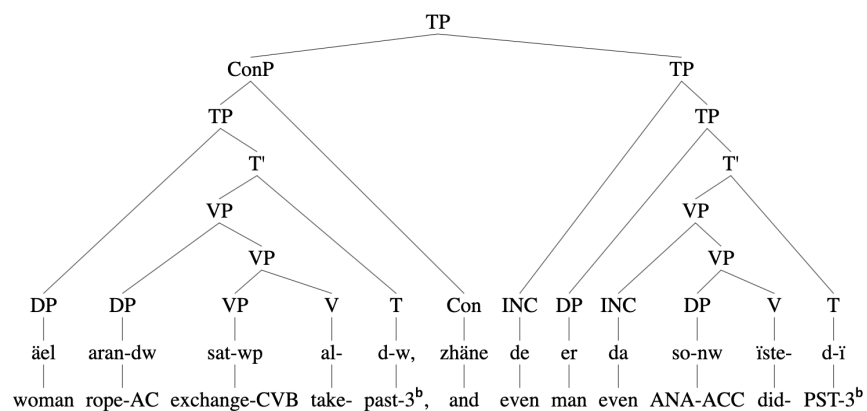
Tense phrases can also be used to select adverb phrases, and can be coordinated with other tense phrases. When they select AdvP's, they are always head-final:

26) “the next day, the child and his puppy had been playing” <F35>, and “it's sunny today” <U1539>:

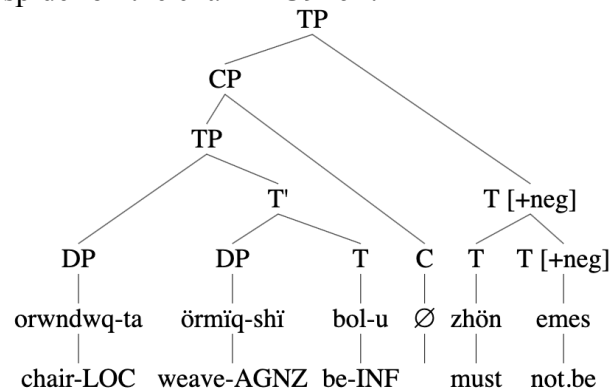


When they select other tense phrases (directly or indirectly), they are also head final:

27) “the woman bought a rope and the man did so too” <U1615>:

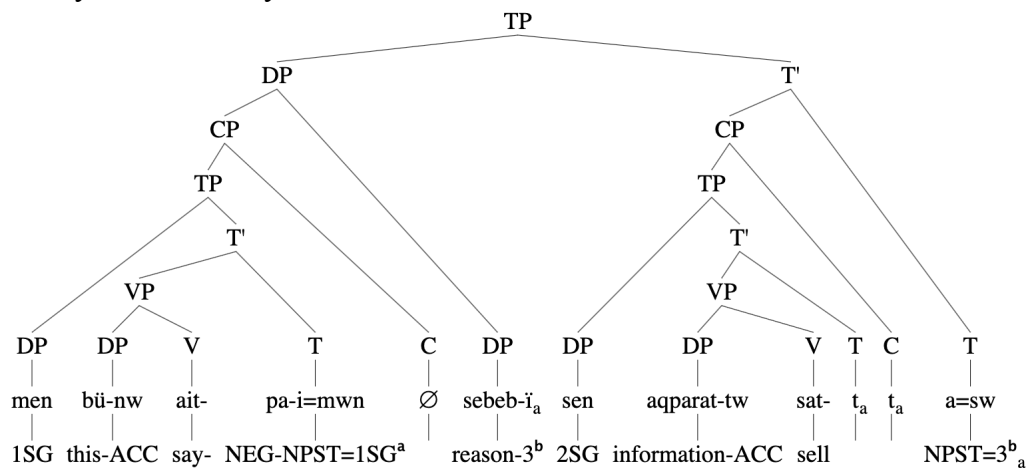


28) “there must not be a spider on the chair” <U946>:



Complementizer phrases are also head-final. In addition to what we've already discussed, consider the following complementizer phrase that demonstrates how the subject “*reason*” moves from the rightmost T position, through the complementizer slot, to the next T position. Because this T position is specified by a VP, the subject can leave its agreement marking. It then works its way up to the specifier of the subject, where it can receive nominal case from the matrix TP:

29) “I don't say that because you sell information.” <U1354>:



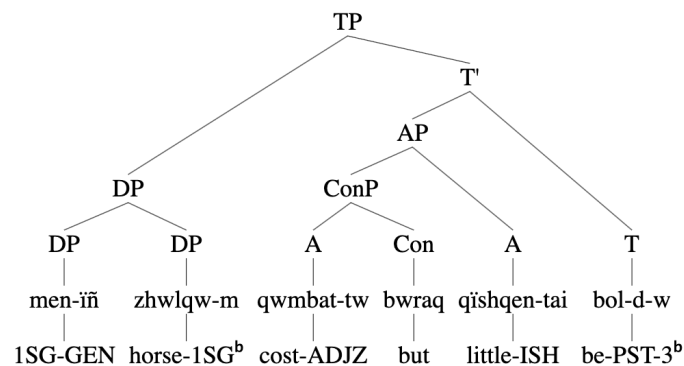
Thus, our model correctly predicts that the grammatical agreement marker at the very end of the above sentence *doesn't* agree with the 2nd person subject of the verb it attaches to, but rather is instead a *third* person marker that agrees with the subject of the matrix tense phrase. This moment also nicely explains why 2PL pronouns can often take the singular form in their agreement markers: The pronominal subject begins in the T head *in the singular form*, marking it for agreement as such, before moving to subject position and being marked for plurality there. Given the earlier discussion of plurals, where we establish a syntactic restriction on plurality marking, it should not be surprising that plural marking should happen after syntactic movement.



The typical selection domain of adpositional heads are other P-elements, and DPs. We see the later above. We expect that we can say things like \*? <tanertern qeiin qyn turalw> to mean something like “*about the day after tomorrow*”, but we have not yet gathered elicitations of this structure. We also predict that these P-elements will pattern with some relativized clauses, as discussed above, once we have collected this data.

Adjectives are never the heads of clauses, unless two adjectives merge to form a count phrase, as explained in the noun syntax section above, or when they are conjoined. In both cases, the phrases are head-final:

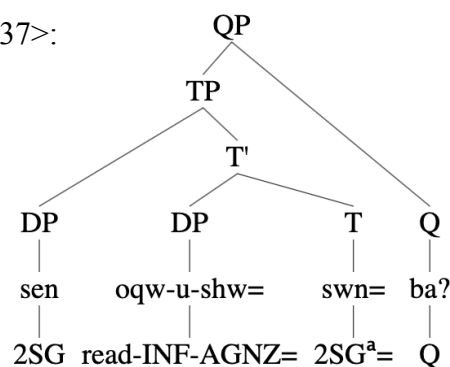
30) “my horse was expensive but small” <U1648>:



## VI. Transformations and Patterns Across Structures

In prior sections, we have already covered most of the many types of subordination and relative clauses generated by Kazakh syntax of which we are already aware. Questions are relatively simple to construct in Kazakh, and are head-final:

31) “are you a student?” <U1637>:



As we have seen, Kazakh constituents are consistently head-final, without known exceptions. Some dependency markings are used, such as between possessor and possessed nouns within noun possession phrases. The [+count] feature is morphologically marked on the head of certain noun phrases, as discussed in section one. Our most ambitious claim about head and dependent markings is the following: All DP's are marked. If the DP is a subject, and thus the dependent of the TP, then it can be marked for nominal case by the subject thematic role, but can never take morphological case. If the DP is not a subject, then it can be marked for morphological case by occupying a specifier role.

## **VII. Conclusion**

This comprehensive review of Kazakh syntax has revealed consistent SOV structure, intricate noun and verb phrase constructions, and a consistent system of head-finality across various syntactic domains. The analysis explains the significance of head-directionality and the interaction between head- and dependent-marking in shaping the syntax of the language. By dissecting noun phrases, verb phrases, and other phrase structures piece by piece and across their entire selection domains, we've been able to examine transformations and patterns across structures.

Many questions regarding Kazakh syntax have been raised throughout this squib. Does Kazakh make functional use of X-phases? Do postpositional phrases pattern, or co-occur, with any relativized DPs? Do morphological postpositions interact with DPs in the ways in which we would expect? Can we find more evidence of T-head subject-role genesis? Does Kazakh have other examples of movement?