

Strategies for the Annotation of Pronominalised Locatives in Turkic Universal Dependency Treebanks

Jonathan Washington¹, Çağrı Çöltekin², Furkan Akkurt³,
Bermet Chontaeva², Soudabeh Eslami², Gulnura Jumalieva⁴, Aida Kasieva⁴,
Aslı Kuzgun, Büşra Marşan⁵, Chihiro Taguchi⁶

¹Swarthmore College, ²University of Tübingen, ³Boğaziçi University, ⁴Kyrgyz-Turkish Manas University,
⁵Stanford University, ⁶University of Notre Dame

MWE-UD Workshop 2024

Attributive *-ki*

	nmod	nsubj	root
Turkish:	<i>oda-da-ki</i>	<i>çocuk-lar</i>	<i>uyu-du-lar</i>
Azerbaijani:	<i>otaq-da-ki</i>	<i>uşaq-lar</i>	<i>yuxla-dı-lar</i>
Kyrgyz:	<i>бөлмө-дө-зү</i>	<i>бал-дар</i>	<i>укта-ды</i>
Tatar:	<i>бүлмә-дә-зе</i>	<i>бала-лар</i>	<i>йокла-ды</i>
Gloss:	room-LOC-ATTR	child-PL	sleep-PST(-PL)
Lemma:	oda/otaq /бөлмө/бүлмә	çocuk/uşaq /бала/бала	uyu/yuxla /укта/йокла
POS:	NOUN	NOUN	VERB
Case:	Loc	Nom	-
Number:	Sing	Plur	Plur

‘The children in the room fell asleep’

-ki/-ki/-зү/-зе suffix allows attributive use of locative nouns

The problem: words with *-ki* can be pronominalized

	amod		nsubj		root
	↙		↘		↓
	<i>büyük</i>		<i>oda-da-ki-ler</i>		<i>uyu-du-lar</i>
Gloss:	big		ROOM-LOC-ATTR-PL		sleep-PST-PL
Lemma:	<i>büyük</i>		<i>oda</i>		<i>uyu</i>
POS:	ADJ		NOUN/PRON		VERB
Case:	-		Loc/Nom		-
Number:	-		Sing/Plur		Plur

'The ones in the big room fell asleep'

The problem: words with *-ki* can be pronominalized

		amod		nsubj		root
		↙		↘		↓
	<i>büyük</i>		<i>oda-da-ki-ler</i>		<i>uyu-du-lar</i>	
Gloss:	big		ROOM-LOC-ATTR-PL		sleep-PST-PL	
Lemma:	<i>büyük</i>		<i>oda</i>		<i>uyu</i>	
POS:	ADJ		NOUN/PRON		VERB	
Case:	-		Loc/Nom		-	
Number:	-		Sing/Plur		Plur	

'The ones in the big room fell asleep'

Subject is the 'the ones in the room', not the 'room'.

The problem: words with *-ki* can be pronominalized

	amod		nsubj		root
	↓		↓		↓
	<i>büyük</i>	<i>oda-da-ki-ler</i>		<i>uyu-du-lar</i>	
Gloss:	big	ROOM-LOC-ATTR-PL		sleep-PST-PL	
Lemma:	büyük	oda		uyu	
POS:	ADJ	NOUN/PRON		VERB	
Case:	-	Loc/Nom		-	
Number:	-	Sing/Plur		Plur	

'The ones in the big room fell asleep'

The adjective modifies the room, not the people in the room.

The problem: words with *-ki* can be pronominalized

	amod		nsubj		root
	↓		↓		↓
	<i>büyük</i>		<i>oda-da-ki-ler</i>		<i>uyu-du-lar</i>
Gloss:	big		ROOM-LOC-ATTR-PL		sleep-PST-PL
Lemma:	büyük		oda		uyu
POS:	ADJ		NOUN/PRON		VERB
Case:	-		Loc/Nom		-
Number:	-		Sing/Plur		Plur

'The ones in the big room fell asleep'

We have multiple, conflicting POS/Number/Case for the same word.

Overview

- The pronominalized locatives violate the lexical integrity principle
- Their treatment across UD Turkic treebanks is not uniform
- Additional issues
 - Multiple *-ki*:
 - *oda-da-ki-nde-ki-ler* ‘The ones with/at the one in the room’ (e.g., referring to plates on the table in the room)
 - *oda-da-ki-nde-ki-ler-de-ki* ‘The one at ones at the one in the room’ (e.g., referring to the food on the plates on the table in the room)
 - ...
 - Pronominalized genitive
oda-nın-ki ‘The one that belongs to the door’
 - Also related to ‘zero-derivation’ from adjectivals to pronominals
büyük-le kes ‘cut with the big (one)’
- We discuss some of the potential/suggested approaches

Obvious solution: multi-word tokens

		aod	nmod	nsubj	root
	<i>büyük</i>		<i>oda-da-ki</i>	<i>-ler</i>	<i>uyu-du-lar</i>
Gloss:	big		room-LOC-ATTR	-PL	sleep-PST-PL
Lemma:	<i>büyük</i>		<i>oda</i>	?	<i>uyu</i>
POS:	ADJ		NOUN	PRON	VERB
Case:	-		Loc	Nom	-
Number:	-		Sing	Plur	Plur

'The ones in the big room fell asleep'

Obvious solution: multi-word tokens – inconsistent analyses

	amod		nmod		nsubj		root	
	<i>büyük</i>	<i>oda-da-ki</i>	<i>-ler</i>	<i>uyu-du-lar</i>				
Gloss:	big	room-LOC-ATTR	-PL	sleep-PST-PL				
Lemma:	<i>büyük</i>	<i>oda</i>	?	<i>uyu</i>				
POS:	ADJ	NOUN	PRON	VERB				
Case:	-	Loc	Nom	-				
Number:	-	Sing	Plur	Plur				

'The ones in the big room fell asleep'

	amod		nsubj		root	
	<i>büyük</i>	<i>oda-da-ki</i>	<i>uyu-du</i>			
Gloss:	big	room-LOC-ATTR	sleep-PST-PL			
Lemma:	<i>büyük</i>	<i>oda</i>	<i>uyu</i>			
POS:	ADJ	NOUN/PRON	VERB			
Case:	-	Loc/Nom	-			
Number:	-	Sing/Sing	Sing			

'The one in the big room fell asleep'

Solutions with no segmentation

		amod		nsubj		root
		↙		↘		↓
	<i>büyük</i>		<i>oda-da-ki-ler</i>		<i>uyu-du-lar</i>	
Gloss:	big		room-LOC-ATTR-PL		sleep-PST-PL	
Lemma:	<i>büyük</i>		<i>oda</i>		<i>uyu</i>	
POS:	ADJ		NOUN/PRON		VERB	
Case:	-		Loc/Nom		-	
Number:	-		Sing/Plur		Plur	

Solutions with no segmentation

+orphan

		amod		orphan		root
		↙		↘		↓
	<i>büyük</i>		<i>oda-da-ki-ler</i>		<i>uyu-du-lar</i>	
Gloss:	big		room-LOC-ATTR-PL		sleep-PST-PL	
Lemma:	<i>büyük</i>		<i>oda</i>		<i>uyu</i>	
POS:	ADJ		NOUN/PRON		VERB	
Case:	-		Loc/Nom		-	
Number:	-		Sing/Plur		Plur	

Solutions with no segmentation

+orphan + 'layered' features

	<i>büyük</i>	<i>oda-da-ki-ler</i>	<i>uyu-du-lar</i>
Gloss:	big	room-LOC-ATTR-PL	sleep-PST-PL
Lemma:	<i>büyük</i>	<i>oda</i>	<i>uyu</i>
POS:	ADJ	NOUN/PRON	VERB
Case:	-	Loc	-
Number:	-	Sing	Plur
Case[2]:	-	Nom	-
Number[2]:	-	Plur	-

```
graph TD
    amod --> bueyuk[büyük]
    orphan --> oda[oda-da-ki-ler]
    root --> uyu[uyu-du-lar]
```

No segmentation: advantages & disadvantages

- + Avoiding (non-trivial) word segmentation
- + No need for empty elements
- + No change to attributive analysis
- The head of the modifier/child is still uncertain
- orphan relation ‘results in a severely under-specified representation’
- Layered/indexed features are unusual/non-standard

Segmentation after *-ki*

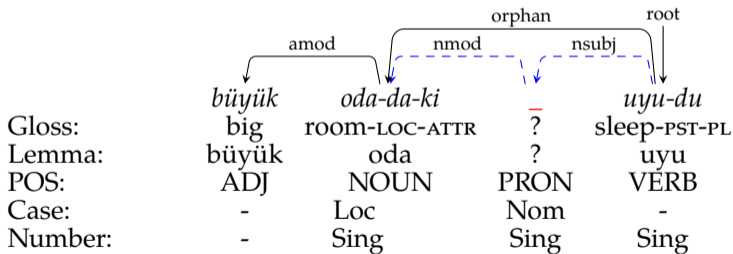
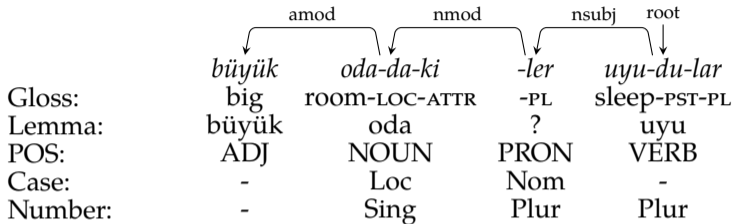
	amod	nmod	nsubj	root
	↙	↘	↙	↘
	<i>büyük</i>	<i>oda-da-ki</i>	<i>-ler</i>	<i>uyu-du-lar</i>
Gloss:	big	room-LOC-ATTR	-PL	sleep-PST-PL
Lemma:	büyük	oda	?	uyu
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Plur	Plur

	amod	nmod	nsubj	root
	↙	↘	↙	↘
	<i>büyük</i>	<i>oda-da-ki</i>	<u>-</u>	<i>uyu-du</i>
Gloss:	big	room-LOC-ATTR	?	sleep-PST-PL
Lemma:	büyük	oda	?	uyu
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Sing	Sing

'Segmenting after': advantages & disadvantages

- Needs segmentation
- Empty forms
- + No change to attributive analysis
- + The head of the modifier/child is clear
- + Modification to the head is clear
- + No conflicting morphological features

Segmentation after *-ki*, fallback to orphan & enhanced dependencies



'Segmenting after' with orphan: advantages & disadvantages

- Needs segmentation
- No empty forms in basic dependencies (reliance on enhanced dependencies)
- + No change to attributive analysis
- + The head of the modifier/child is clear in basic dependencies
- orphan relation basic dependencies
- Possibly conflicting morphological features

Segmentation before *-ki*

	amod	nmod	nsubj	root
	↙	↘	↙	↘
	<i>büyük</i>	<i>oda-da</i>	<i>-ki-ler</i>	<i>uyu-du-lar</i>
Gloss:	big	room-LOC-ATTR	-ATTR-PL	sleep-PST-PL
Lemma:	<i>büyük</i>	<i>oda</i>	<i>-ki</i>	<i>uyu</i>
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Plur	Plur

	amod	nmod	nsubj	root
	↙	↘	↙	↘
	<i>büyük</i>	<i>oda-da</i>	<i>-ki</i>	<i>uyu-du</i>
Gloss:	big	room-LOC-ATTR	-ATTR	sleep-PST-PL
Lemma:	<i>büyük</i>	<i>oda</i>	<i>-ki</i>	<i>uyu</i>
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Sing	Sing

'Segmenting before': advantages & disadvantages

- Needs segmentation (maybe also for attributive)
- + No empty forms
- Attributive version may also need segmentation
- + The head of the modifier/child is clear
- + Modification to the head is clear
- + No conflicting morphological features

A reluctant proposal: segment before

	amod	nmod	nsubj	root
	<i>büyük</i>	<i>oda-da</i>	<i>-ki(-ler)</i>	<i>uyu-du</i>
Gloss:	big	room-LOC	-ATTR(-PL)	sleep-PST-PL
Lemma:	<i>büyük</i>	<i>oda</i>	<i>-ki</i>	<i>uyu</i>
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Sing(Plur)	Sing

A reluctant proposal: segment before

	amod	nmod	nsubj	root
	↓	↓	↓	↓
	<i>büyük</i>	<i>oda-da</i>	<i>-ki(-ler)</i>	<i>uyu-du</i>
Gloss:	big	room-LOC	-ATTR(-PL)	sleep-PST-PL
Lemma:	büyük	oda	-ki	uyu
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Sing(Plur)	Sing

	amod	nmod	nsubj	root
	↓	↓	↓	↓
	<i>büyük</i>	<i>oda-da</i>	<i>-ki</i>	<i>çocuklar</i>
Gloss:	big	room-LOC	-ATTR	children
Lemma:	büyük	oda	-ki	çocuk
POS:	ADJ	NOUN	ADP	NOUN
Case:	-	Loc	-	Nom
Number:	-	Sing	-	Plur

A reluctant proposal: segment before

	amod	nmod	nsubj	root
	büyük	oda-da	-ki(-ler)	uyu-du
Gloss:	big	room-LOC	-ATTR(-PL)	sleep-PST-PL
Lemma:	büyük	oda	-ki	uyu
POS:	ADJ	NOUN	PRON	VERB
Case:	-	Loc	Nom	-
Number:	-	Sing	Sing(Plur)	Sing

	amod	nmod	nsubj	root
	büyük	oda-da	-ki	çocuklar
Gloss:	big	room-LOC	-ATTR	children
Lemma:	büyük	oda	-ki	çocuk
POS:	ADJ	NOUN	ADP	NOUN
Case:	-	Loc	-	Nom
Number:	-	Sing	-	Plur

	amod	case/dep	nsubj	root
	büyük	oda-da	-ki	çocuklar
Gloss:	big	room-LOC	-ATTR	children
Lemma:	büyük	oda	-ki	çocuk
POS:	ADJ	NOUN	ADP	NOUN
Case:	-	Loc	-	Nom
Number:	-	Sing	-	Plur

	amod	case/dep	nsubj	root
	büyük	oda-da	-ki	uyu-du
Gloss:	big	room-LOC	-ATTR	sleep-PST-PL
Lemma:	büyük	oda	-ki	uyu
POS:	ADJ	NOUN	ADP	VERB
Case:	-	Loc	-	-
Number:	-	Sing	-	Sing

Thank you for listening! Questions/suggestions are very welcome.

Locative noun without attributive

	obl		root
		nsubj	
Turkish:	<i>oda-da</i>	<i>çocuk-lar</i>	<i>uyu-du-lar</i>
Azerbaijani:	<i>otaq-da</i>	<i>uşaq-lar</i>	<i>yuxla-dı-lar</i>
Kyrgyz:	<i>бөлмө-дө</i>	<i>бал-дар</i>	<i>укта-ды</i>
Tatar:	<i>бүлмә-дә</i>	<i>бала-лар</i>	<i>йокла-ды</i>
	room-LOC	child-PL	sleep-PST(-PL)
Lemma:	<i>oda/otaq</i> <i>/бөлмө/бүлмә</i>	<i>çocuk/uşaq</i> <i>/бала/бала</i>	<i>uyu/yuxla</i> <i>/укта/йокла</i>
POS:	NOUN	NOUN	VERB
Case:	Loc	Nom	-

'Children slept in the room.'

Longer example, including possessive

		nmod:poss		nmod:loc		nsubj	root
	<i>Bardak</i>		<i>dolab-in-da-ki</i>		<i>bardak-lar-ım</i>		<i>düş-tü-ler</i>
	cup		cabinet-POSS.3-LOC-ATTR		cup-PL-POSS.1SG		fall-PST-3PL
Lemma:	bardak		dolap		bardak		düş
POS:	NOUN		NOUN		NOUN		VERB
Number:	Sing		Sing		Plur		Plur
Person[psor]:	-		3		1		-
Case:	Nom		Loc		Nom		-

'My glasses in the cup cabinet fell'

No segmentation

		nmod:poss		nsubj/orphan		root
		↓		↓		↓
	<i>Bardak</i>		<i>dolab-in-da-ki-ler-im</i>		<i>düş-tü-ler</i>	
	cup		cabinet-POSS.3-LOC-ATTR-PL-POSS.1SG		fall-PST-3PL	
Lemma:	bardak		dolap		düş	
POS:	NOUN		NOUN		VERB	
Number:	Sing		Sing/Plur		Plur	
Person[psor]:	-		3/1		-	
Case:	Nom		Loc/Nom		-	

'The ones of mine on the cup cabinet fell'

No segmentation, using (extended) layered features.

	<i>Bardak</i>	<i>dolab-ın-da-ki-ler-im</i>	<i>düş-tü-ler</i>
	cup	cabinet-POSS.3-LOC-ATTR-PL-POSS.1SG	fall-PST-3PL
Lemma:	bardak	dolap	düş
POS:	NOUN	NOUN	VERB
Number[stem1]:	-	Sing	-
Number[stem2]:	-	Plur	-
Person[psor][stem1]:	-	3	-
Person[psor][stem2]:	-	1	-
Case[stem1]:	-	Loc	-
Case[stem2]:	-	Nom	-

Diagram labels: *nmod:poss* (above Bardak), *nsubj/orphan* (above dolab-ın-da-ki-ler-im), *root* (above düş-tü-ler)

