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

Jonathan Washington\textsuperscript{1}, Çağrı Çöltekin\textsuperscript{2}, Furkan Akkurt\textsuperscript{3}, Bermet Chontaeva\textsuperscript{2}, Soudabeh Eslami\textsuperscript{2}, Gulnura Jumalieva\textsuperscript{4}, Aida Kasieva\textsuperscript{4}, Aslı Kuzgun, Büşra Marşan\textsuperscript{5}, Chihiro Taguchi\textsuperscript{6}

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

MWE-UD Workshop 2024
## Attributive -ki

<table>
<thead>
<tr>
<th>Turkish:</th>
<th>oda-da-ki</th>
<th>çocuk-lar</th>
<th>uyu-du-lar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azerbaijani:</td>
<td>otaq-da-ki</td>
<td>uşaq-lar</td>
<td>yuxla-di-lar</td>
</tr>
<tr>
<td>Kyrgyz:</td>
<td>бөлөмө-дө-гү</td>
<td>бал-дар</td>
<td>укта-ды</td>
</tr>
<tr>
<td>Tatar:</td>
<td>булмә-дә-ге</td>
<td>бала-лар</td>
<td>йокла-ды</td>
</tr>
</tbody>
</table>

### Gloss
- room-LOC-ATTR
- child-PL
- sleep-PST(-PL)

### Lemma
- oda/otaq
- çocuk/uşaq
- uyu/yuxla

### POS
- NOUN
- NOUN
- VERB

### Case
- Loc
- Nom
- -

### Number
- Sing
- Plur
- Plur

'The children in the room fell asleep'

-ki/-ki/-гү/-ге suffix allows attributive use of locative nouns
The problem: words with -\textit{ki} can be pronominalized

\begin{itemize}
\item \textit{büyük} (big)
\item \textit{oda-da-ki-ler} (room-LOC-ATTR-PL)
\item \textit{uyu-du-lar} (sleep-PST-PL)
\end{itemize}

\textbf{Gloss:} \textit{büyük} \textit{oda-da-ki-ler} \textit{uyu-du-lar}
\textbf{Lemma:} \textit{büyük} \textit{oda} \textit{uyu}
\textbf{POS:} \textbf{ADJ} \textbf{NOUN/PRON} \textbf{VERB}
\textbf{Case:} - \textbf{Loc/Nom} -
\textbf{Number:} - \textbf{Sing/Plur} \textbf{Plur}

\begin{itemize}
\item \textit{The ones in the big room fell asleep’}
\end{itemize}
The problem: words with -ki can be pronominalized

Subject is the ‘the ones in the room’, not the ‘room’.
The problem: words with -

-ki can be pronominalized

-ki

büyük oda-da-ki-ler uyu-du-ler

big room-LOC-ATTR-PL sleep-PST-PL

ADJ NOUN/PRON VERB

- Loc/Nom -

Sing/Plur Plur

Gloss: Lemma: POS: Case: Number:

‘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

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

‘The ones in the big room fell asleep’
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-nin-ki ‘The one that belongs to the door’
  - Also related to ‘zero-derivation’ from adjectival to pronouns
    büyük-le kes ‘cut with the big (one)’
- We discuss some of the potential/suggested approaches
Obvious solution: multi-word tokens

Gloss:  
Lemma:  
POS:  
Case:  
Number:  

‘The ones in the big room fell asleep’
Obvious solution: multi-word tokens – inconsistent analyses

‘The ones in the big room fell asleep’

‘The one in the big room fell asleep’
Solutions with no segmentation

Gloss: büyük oda-da-ki-lar uyuy-du-lar
Lemma: büyük oda uyuy
POS: ADJ NOUN/PRON VERB
Case: - Loc/Nom -
Number: - Sing/Plur Plur
Solutions with no segmentation

+orphan

Gloss:  büyük  oda-da-ki-ler  uyu-du-lar
Lemma: büyük  oda  uyu
POS:  ADJ  NOUN/PRON  VERB
Case: -  Loc/Nom  -
Number: -  Sing/Plur  Plur
Solutions with no segmentation
+orphan + ‘layered’ features

Gloss: 
Lemma: 
POS: 
Case: 
Number: 
Case[2]: 
Number[2]:

buyük big oda-da-ki-ler room-LOC-ATTR-PL uyuu-du-lar sleep-PST-PL
ADJ NOUN/PRON VERB
Loc Sing Plur Nom
- - -
- - Plur
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
  – orphant relation ‘results in a severely under-specified representation’
  – Layered/indexed features are unusual/non-standard
Segmentation after -*ki*

**Gloss:**
büyük oda-da-ki -ler uyu-du-lar

**Lemma:**
büyük oda da-ki -ler uyu-du-lar

**POS:**
ADJ NOUN PRON VERB

**Case:**
Loc Nom -

**Number:**
Sing Plur Plur

---

**Gloss:**
büyük oda-da-ki uyu-du

**Lemma:**
büyük oda da-ki uyu-du

**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

Gloss: büyük oda-da-ki -ler uyu-du-lar
Lemma: büyük oda -ler uyu
POS: ADJ NOUN PRON VERB
Case: - Loc Nom -
Number: - Sing Plur -

Gloss: büyük oda-da-ki uyu-du
Lemma: büyük oda -ler uyu
POS: ADJ NOUN PRON VERB
Case: - Loc Nom -
Number: - Sing Sing Sing
‘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

Gloss:   büyük  oda-da  -ki-ler  uyu-du-lar
Lemma:  büyük  oda  -ki  uyu
POS:  ADJ  NOUN  PRON  VERB
Case:  -  Loc  Nom  -
Number:  -  Sing  Plur  Plur

Gloss:   büyük  oda-da  -ki  uyu-du
Lemma:  büyük  oda  -ki  uyu
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

Gloss:   big   room-LOC   -ATTR(-PL)   sleep-PST-PL
Lemma:  büyük   oda   -ki   uyu
POS:  ADJ   NOUN   PRON   VÉRB
Case:   -   Loc   Nom   -
Number:  -   Sing   Sing(Plur)   Sing
A reluctant proposal: segment before

Thank you for listening! Questions/suggestions are very welcome.
A reluctant proposal: segment before

Thank you for listening! Questions/suggestions are very welcome.
Locative noun without attributive

Turkish: oda-da çocuk-lar uydu-lar
Azerbaijani: otaq-da uşaq-lar yuxla-dı-lar
Kyrgyz: бөлөө-де балдар уктады
Tatar: булма-де бала-лар укта-ды

Lemma: oda/otaq çocuk/uşaq uyу/yuxla
POS: NOUN NOUN VERB
Case: Loc Nom -

‘Children slept in the room.’
Longer example, including possessive

My glasses in the cup cabinet fell’
The ones of mine on the cup cabinet fell.
No segmentation, using (extended) layered features.

<table>
<thead>
<tr>
<th>Lemma</th>
<th>POS</th>
<th>Number[stem1]:</th>
<th>Number[stem2]:</th>
<th>Person[psor][stem1]:</th>
<th>Person[psor][stem2]:</th>
<th>Case[stem1]:</th>
<th>Case[stem2]:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bardak</td>
<td>NOUN</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>cup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dolab-ın-da-ki-ler-im</td>
<td>NOUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cabinet-POSS.3-LOC-ATTR-PL-POSS.1SG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>düş-tü-ler</td>
<td>VERB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>düş</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

UD-Turkic Group  Turkic Pronominalized Locatives  MWE-UD Workshop 2024  A.4