ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

QARAQALPAQ TILI USHÍN SINTAKSISLIK ANALIZ TÚRLERINIŃ SALÍSTÍRMALÍ ANALIZI

Mambetullaeva Biybihajar
Qaraqalpaq Mámleketlik Universiteti 2-kurs magistrantı.
https://doi.org/10.5281/zenodo.19384707

Annotaciya. *Sintaksislik analiz tábiyiy tildi qayta islewdiń tiykarǵı basqıshlarınan biri bolıp, gáptiń grammatikalıq dúzilisin ashıp beriwge xızmet etedi. Búgingi kúnde birqansha sintaksislik analiz túrleri bar bolıp, olardan constituency hám dependency parsing túrleri keńnen qollanıp kelinbekte. Jumıstıń maqseti bul eki analiz túrin salıstırıw hám qaraqalpaq tili grammatikasın ashıp beriwdegi abzallıqların bahalaw bolıp tabıladı. Maqalada parsing túrlerin, olardıń bir-birinen parqın analizlew hám qaraqalpaq tiliniń sintaksislik analizatorın jaratıwda eń maqul bolǵan joldı kórsetiwge háreket etildi.*

Tayanısh túsinikler: *sintaksislik analiz, dependency parsing, constituency parsing, NLP, agglyutinativ til.*

Tábiyiy tildi qayta islewdiń rawajlanıwı menen informaciyalıq sistemalardı, atap aytqanda, mashina awdarması, avtomatikalıq analiz, tekst analizi hám dialog sistemaların jaratıw jedellesip barmaqta. Tekstti qayta islewdiń áhmiyetli basqıshlarınan biri gáp strukturasın anıqlaw hám sózler arasında grammatikalıq baylanıslardı tabıwǵa qaratılǵan sintaksislik analiz bolıp tabıladı.

Hárqanday tábiyiy til óziniń sintaksislik nızamlıqlarına iye. Sol nızamlıqlar arqalı sózler bir-biri menen baylanısadı hám sóz dizbegin, gáp túrlerin payda etedi. Bunday sintaksislik baylanıslardı parser arqalı analizlewdiń birneshe túr hám metodları bar bolıp, olar búgingi kúnde kóplegen tillerdiń sintaksislik analizatorların jaratıwda keńnen qollanıp kelinbekte. Sintaksislik parsing túrleri boyınsha ámelge asırılǵan izertlew jumısların talqılaw barısında Constituent Parsing, Dependency Parsing, Semantic Dependency Graph, Cross-Domain Parsing, Cross-Lingual Parsing [3] sıyaqlı analiz túrleri boyınsha úyreniw múmkin. Bul parsing túrleri aldınnan tayarlanǵan úlken kólemdegi korpuslardıń jaratılıwı hám neyron tarmaqlardıń qosılıwı esabınan kóplegen tillerge endirilip, jedel túrde rawajlanıwına qaramastan, az resurslı tiller óz sheshimin kútip turǵan áhmiyetli máselelerdiń biri bolıp qalmaqta. Sonlıqtan maqalamızda tiykarınan constituency hám dependency parsing túrlerin hám olardıń agglyuntinativ tiller qatarına kiriwshi túrkiy tiller, sonıń ishinde qaraqalpaq tiliniń sintaksislik dúzilisin ashıp beriwi máselesin salıstırmalı túrde talqılaymız.

Jumısta kompyuter lingvistikası hám tábiyiy tildi qayta islew tarawındaǵı ilimiy ádebiyatlardı analizlew hám ulıwmalastırıw usıllarınan paydalanılǵan. Izertlew materialı sıpatında sintaksislik analizator tarawına arnalǵan jumıslar, ilimiy materiallar, sonıń ishinde xalıqaralıq konferenciyalar materialları hám ashıq ilimiy bazalardaǵı resurslardan, sonday-aq, házirgi kúnde iske qosılǵan túrkiy til parserlerinen paydalandıq.

Constituency parsing gáptegi izbe-iz jaylasqan jaylasqan sintaksislik toparlardı ierarxiyalıq dúzilis kórinisinde beriwge tiykarlanǵan. Bunda parserdiń wazıypası gáp qurılısın analizlew bolıp, gáp (S) eń dáslep baslawısh basqarıwındaǵı sózler (NP) hám bayanlawısh basqarıwındaǵı sózlerge (VP) ajıratıp alınadı [4]. Yaǵnıy, bul parsing túri gápti quram bóleklerge ajıratadı.

ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

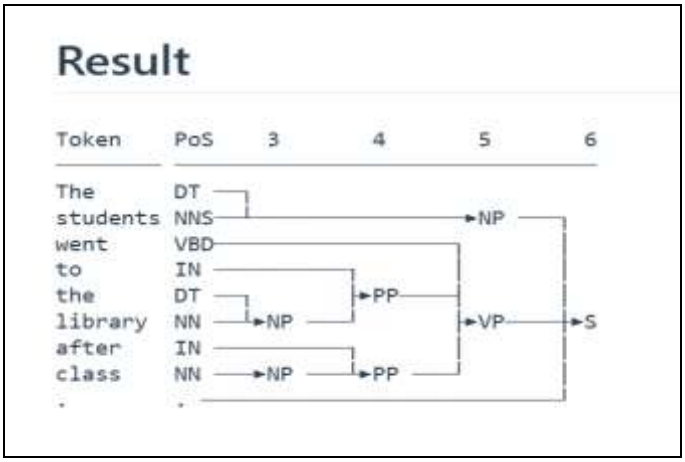
Analiz nátiyjesi gáptiń quramlıq bólimlerge bóliniwin sáwlelendiriwshi terek bolıp, onı dúziw ushın grammatikalıq qaǵıydalar toplamı hám grammatikalıq modellerden paydalanadı.

Constituency parsingniń maqseti sóz dizbeklerin hám olardıń baylanısın ashıp beriwden ibarat. Bul parsing túri búgingi kúnde tekstti analizlew, mashina awdarması, sorawlarǵa juwap beriw, tekstti klassifikaciyalaw, informaciya izlew sıyaqlı túrli NLP wazıypalarında qollanılmaqta.

Házirgi waqıtta agglyutinativ tiller ushın constituency parserler derlik qollanılmaydı.

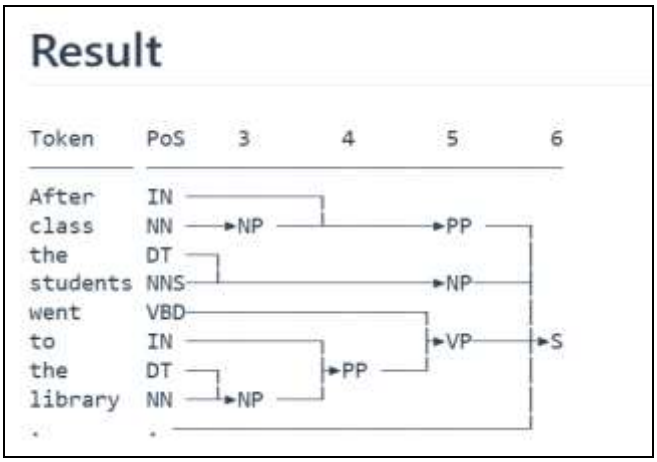
Bunıń sebebi constituency parsing kóbirek turaqlı orın tártipke iye sintaksislik dúzilistegi gáplerdi analizlewge baǵdarlanǵan bolıp, agglyutinativ tiller, sol qatarda qaraqalpaq tiline tán erkin orın tártip hám de uzın gáplerdi analizlewde qáte-kemshiliklerge jol qoyıwı múmkin.

Qaraqalpaq tilinde tayar parser járdeminde analizlew imkaniyatı bolmaǵanlıqtan, inglis tilindegi gápti analizlew arqalı bul analiz túriniń ózgesheligin ashıp beriwge háreket ettik. Dáslep HanLP (https://hanlp.hankcs.com/en/demos/con.html) saytına input, yaǵnıy «The students went to the library after class» hám tómendegi nátiyjeni aldıq:



Mısalda kórinip turǵanınday, gáp óz izbe-izligi boyınsha tokenlerge ajıratılǵan, soń morfologiyalıq jaqtan teglengen. Sintaksislik jaqtan gáp tolıǵı menen S tegi menen belgilengen. S eki quram bólekke - NP hám VP teglerine ajıratılǵan, VP teginiń ózi eki pısıqlawıshtı (PP) ózinde birlestirgenin kóremiz.

Mısaldı qaraqalpaq tiline awdarmalaw arqalı parsingniń imkaniyatların jaqsıraq analizlew múmkin: [S [NP Studentler] [VP [AdvP sabaqtan soń] [PP kitapxanaǵa] [V bardı]].]. Endi tap usı gápti inversiyalıq orın tártipte bereyik:



ISSN: 2181-3906 2026

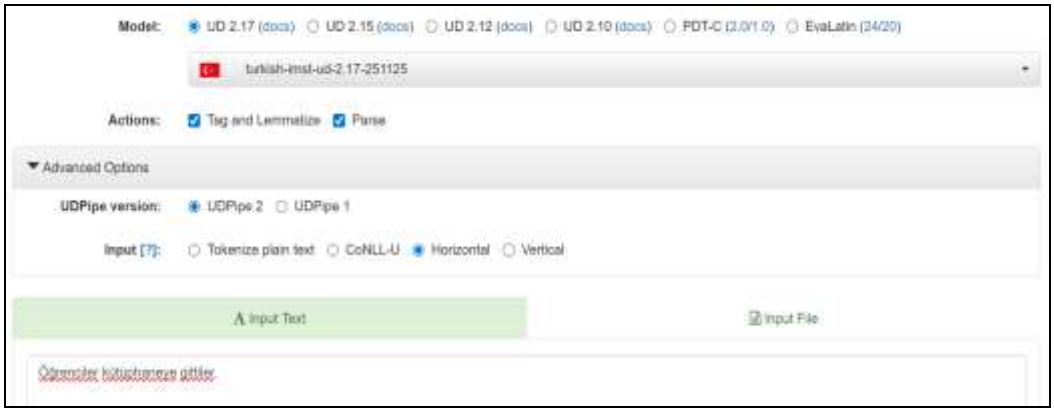
International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

Bunda gáp (S) úsh quramlıq bólekke (PP, NP, VP) ajıratılǵan. Bul nátiyjeni de qaraqalpaq tili mısalında talqılayıq: [S [AdvP sabaqtan soń] [NP studentler] [VP [PP kitapxanaǵa] [V bardı]].]. Kórinip turǵanınday, baslawıshtan aldın kelgen pısıqlawıshtıń gáp aqırındaǵı bayanlawısh penen baylanısı ashıp berilmegen. Demek, parser qáteliklerge jol qoymawı ushın hárbir gáp aǵzası tuwrı orın tártipte qollanılıwı kerek.

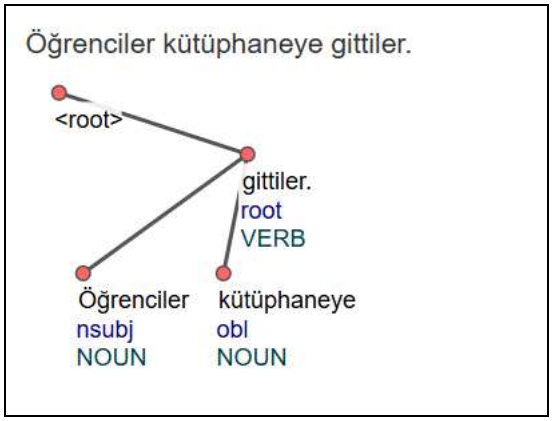
Dependency parsing constituency parsingniń qarama-qarsı túri bolıp, ol gápti quram bóleklerge ajıratpastan, gáp aǵzaları arasındaǵı baǵınıńqı baylanıstı ashıp beredi. Bayanlawısh (Root) gáptiń tiykarı retinde qaraladı. Biraq gáptegi barlıq aǵzalardı rootqa tikkeley baylanıstırıp qoymastan, sózlerdiń arasındaǵı baylanısta baǵınıńqı hám bas sózdi kórsetip beredi. Bul jaǵınan dependency parsing agglyutinativ tillerdiń grammatikasına say keledi. Sebebi, ol «bas – baǵınıńqı» qatnas kórinisindegi sintaksislik funkcional qurılıstı ózinde sáwlelendiredi. Onıń nátiyjesi gáp aǵzaları arasındaǵı baǵınıńqı baylanıstı kórsetiwshi baǵınıńqılıq teregi (dependency tree) bolıp tabıladı.

Házirgi kúndegi zamanagóy NLP sistemalarında dependency parsing tekstti tanıw, teglew, tekst generaciyasında keń qollanılmaqta. Iri kólemli joybarlardıń biri bolǵan Univesal Dependencies te tiykarınan usı analiz túrinen paydalanadı.

Bul parsing túriniń ózgesheligin kórsetiw ushın qaraqalpaq tiline sintaksislik qurılısı jaǵınan uqsas bolǵan túrk tilindegi gápti Lindat (https://lindat.mff.cuni.cz/services/udpipe/) platformasında analizledik. Dáslep Turkish-IMST modelin hám UDPipe kitapxanasınıń 2.7 variantın tańlaymız. Input «Studentler kitapxanaǵa ketti»:



Nátiyje:

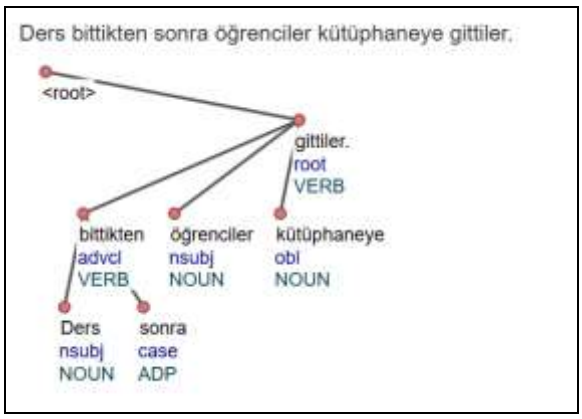


Bunda «gittiler» sózi gáptiń tiykarı sıpatında kórsetilgen.

ISSN: 2181-3906 2026

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 5 / ISSUE 3 / UIF:8.2 / MODERNSCIENCE.UZ

Gápti quramastırǵanda, tómendegi nátiyjeni alamız:



1	Ders	ders	NOUN	Noun	Case=Nom|Number=Sing|Person=3	2	nsubj
2	bittikten	bit	VERB	Verb	Aspect=Perf|Case=Abl|Mood=Ind|Polarity=Pos|Tense=Past|VerbForm=Part	6	advcl
3	sonra	sonra	ADP	PCAbl	_	2	case
4	öğrenciler	öğrenci	NOUN	Noun	Case=Nom|Number=Plur|Person=3	6	nsubj
5	kütüphaneye	kütüphane	NOUN	Noun	Case=Dat|Number=Sing|Person=3	6	obl
6	gittiler.	gittiler	VERB	Verb	_	0	root

Gápte baǵınıńqı baylanıslar durıs analizlengen. Qospa gáptiń bas gápindegi bayanlawısh gáptiń tiykarı retinde alınǵan. Sonday-aq, baǵınıńqı aǵzanıń qaysı aǵzaǵa baylanısqanı Head qatarında, onnan sońǵı qatarda gáp aǵzaları kórsetilgen: «Ders» - nsubj (baslawısh), «bittikten» - advcl (pısıqlawısh baǵınıńqı gáp), «sonra» - case (tirkewish), «öğrenciler» — nsubj (baslawısh), «kütüphaneye» — obl (pısıqlawısh), «gittiler» — root (bayanlawısh). Dependency treede gápler óz orın tártibinde emes, al baǵınıńqılıq ierarxiyası tiykarında tártiplestirilgenin kóriwimiz múmkin. Tap usı parsing túri arqalı qaraqalpaq tilindegi gáp analizlew imkaniyatı bolǵanda, tómendegi nátiyjeni alıw múmkin:

1. Sabaq — noun — 2 — nsubj
2. juwmaqlanǵannan — verb — 6 — advcl
3. soń — adp — 2 — case
4. studentler — noun — 6 — nsubj
5. kitapxanaǵa — noun — 6 — obl
6. ketti — verb — 0 — root

Mısalda kórinip turǵanınday, bul analiz túri qaraqalpaq tili sintaksislik qaǵıydalarına ádewir jaqın keliwi hám qáte-kemshiliklersiz tallay alıwı jaǵınan salıstırmalı dárejede abzallıqqa iye.

Joqarıda analizlengen eki parsing túri de házirgi NLP wazıypaların ámelge asırıwda óz ornına iye. Constituency parsing gáptiń quram bóleklerin kórsetiw, gáp toparların izbe-izlikte terek kórinisinde vizuallastırıw imkaniyatın berse, dependency parsing arqalı sózler ortasındaǵı baǵınıqı baylanıstı anıqlaw, bas aǵzalardı kórsetiw múmkin.

Izertlewden alınǵan nátiyjeler sintaksislik analiz túrlerin tańlaw tikkeley tildiń grammatikalıq ózgesheliklerine baylanıslı ekenligin kórsetedi. Turaqlı orın tártip (SVO) hám salıstırmalı ápiwayı morfologiyaǵa iye analitikalıq tillerdiń sintaksislik ózgesheliklerin ashıp beriwdi constituency parsing arqalı orınlaw hám anıqlıq, hám sistemalılıqtı beredi. Biraq, sózler arasındaǵı baylanıstı támiyinlew orın tártip emes, al kóp sanlı morfologiyalıq kategoriyalar arqalı támiyinlenetuǵın agglyutinativ (SOV) tiller ushın dependency parsingti qollanıw eń durıs hám nátiyjeli hám jol esaplanadı.

Degen menen qaraqalpaq tiliniń sintaksislik analizatorın jaratıwda durıs tańlanǵan parsing túrin qollanıw jetkiliksiz.

Parser tildiń grammatikasın qátesiz kórsetiwi ushın, ol sol tildiń grammatikalıq qaǵıydaları tiykarında islewi, yaǵnıy grammatikalıq nızamlıqlarǵa úyretilgen bolıwı kerek. Sonıń menen birge úlken kólemli korpus jaratıw, ilajı barınsha kóp sandaǵı gáplerdi annotaciyalaw analizatordıń durıs islewin támiyinleydi.

Paydalanılǵan ádebiyatlar:

1. Dáwletov A, Dáwletov M, Qudaybergenov M. Házirgi qaraqalpaq ádebiy tili. Sintaksis. Nókis – 2009.
2. Zong Chengqing. Statistical Natural Language Processing. - Tsinghua University Press, 2013.
3. Meishan Zhang. A Survey of Syntactic-Semantic Parsing Based on Constituent and Dependency Structures. Science China Press and Springer -Verlag Berlin Heidelberg, 2017.
4. Abjalova M, Tahrir va tahlil dasturlarining lingvistik modullari. Toshkent – 2020.
5. Бегфорт Б, Билбро Р, Охеда Т. Прикладной анализ текстовых данных на Python. Санкт-Петербург – 2019.
6. Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd (draft) ed.). Prentice Hall PTR, 2019.
7. https://www.scaler.com/topics/nlp/introduction-to-grammar-in-nlp/