T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

Özbekçe'nin Biçimbilimsel ve Sözdizimsel Çözümlemesi ve Özbekçe-Türkçe Makine Çevirisi Sistemi

Ahmet Eralp KÜÇÜK

DANIŞMAN
Doç. Dr. Atakan KURT

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

İSTANBUL-2022

Bu çalışma, [18.02.2022] tarihinde aşağıdaki jüri tarafından [Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Programı]nda [Doktora tezi] olarak kabul edilmiştir.

Tez Jürisi

[Doç. Dr.] [Atakan KURT](Danışman)
[İstanbul Üniversitesi-Cerrahpaşa]
[Mühendislik Fakültesi]

[Prof. Dr.] [Ahmet SERTBAŞ]
[İstanbul Üniversitesi-Cerrahpaşa]
[Mühendislik Fakültesi]

[Prof. Dr.] [Erhan AKIN]
[Fırat Üniversitesi]
[Mühendislik Fakültesi]

[Doç. Dr.] [Fatih KOÇAN]
[Atlas Üniversitesi]
[Mühendislik ve Doğa Bilimleri Fakültesi]

[Dr. Öğr. Üyesi] [Muhammed Erdem İSENKUL]
[İstanbul Üniversitesi-Cerrahpaşa]
[Mühendislik Fakültesi]

20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi-Cerrahpaşa'nın abonesi olduğu intihal yazılım programı kullanılarak Lisansüstü Eğitim Enstitüsü'nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

ÖNSÖZ

Çalışmaları ile öğrencilerine ışık tutan, lisans eğitimimden doktora eğitimime ve bu tezin ortaya çıkması için fikir, görüş, bilgi ve akademik tecrübelerinden yararlandığım saygıdeğer hocam sayın Doç. Dr. Atakan KURT'a teşekkürü bir borç bilirim.

Bilgi, öneri ve yardımlarıyla tez çalışmama önemli katkılar sağlayan saygıdeğer jüri üyeleri Prof. Dr. Ahmet SERTBAŞ, Prof. Dr. Erhan AKIN ve tez çalışmam esnasında ebediyete intikal eden Doç. Dr. Veli HAKKOYMAZ'a teşekkürlerimi sunarım.

Dili ve dilbilgisini öğrenmemde büyük katkıları olan Dr. Kalmamat KULAMSHAEV, tez uygulamamda her türlü yardımda bulunan Harun Reşit ZAFER'e ve çalışmalarım sırasında benden hiçbir desteği esirgemeyen sevgili annem ve babam, Engin ve Mehmet KÜÇÜK'e, kardeşlerim Fethi Namık KÜÇÜK ve Sidre KÜÇÜK KARAASLAN ile yol arkadaşım, sevgili eşim Esra KÜÇÜK'e sonsuz teşekkür ederim.

Yüzyıllar öncesinden Türk diline katkı sağlayan ve çalışmalarını örnek aldığımız başta Kaşgârlı Mahmud olmak üzere tüm edebiyat, bilim insanlarına ve tüm zorluklara rağmen büyük bir fedakarlıkla ülkesini geliştirmek için çabalayan, emek harcayan, çalışmalarımı emanet edebileceğim ve bu çalışmaları geliştireceklerini ümit ettiğim idealist Türk Gençlerine ithaf olunmuştur.

Şubat 2022 Ahmet Eralp KÜÇÜK

İÇİNDEKİLER

Sayfa No

ÖNSÖZ ..... **iv**
İÇİNDEKİLER ..... **v**
ŞEKİL LİSTESİ ..... **viii**
TABLO LİSTESİ ..... **x**
SİMGE VE KISALTMA LİSTESİ ..... **xi**
ÖZET ..... **xii**
SUMMARY ..... **xiii**
1. GİRİŞ **1**
1.1. DOĞAL DİL İŞLEME ÇALIŞMALARI 2
1.2. TERMİNOLOJİ 3
1.3. PROBLEM TANIMI VE ÖNERİLEN ÇÖZÜM 4
1.4. İLGİLİ ÇALIŞMALAR 6
2. GENEL KISIMLAR **10**
2.1. ÖZBEK DİLİNE GENEL BAKIŞ VE ÖZBEKÇE GENEL BİLGİLER 10
2.2. ÖZBEKÇE SES BİLGİSİ, SES OLAYLARI VE SES DEĞİŞİMLERİ 12
2.2.1. Hece ve Hece Türleri 12
2.2.2. Vurgu 13
2.2.3. Ses Uyumları 13
2.2.3.1. Büyük Ünlü Uyumu *14*
2.2.3.2. Küçük Ünlü Uyumu *14*
2.2.3.3. Ünsüz Uyumu *15*
2.3. ÖZBEKÇE YAPIM VE ÇEKİM EKLERİ 15
2.3.1. Kodlama Şeması 16
2.3.2. İsim Çekim Ekleri 17
2.3.2.1. Çokluk Eki *21*
2.3.2.2. İyelik Ekleri *21*
2.3.2.3. İsim Hâl Ekleri *22*
2.3.2.4. Soru Eki *23*
2.3.2.5. Olumsuzluk Eki *23*
2.3.3. İsim Yapım Ekleri 23

2.3.3.1. İsimden İsim Yapan Yapım Ekleri *24*
2.3.3.2. Sıfattan İsim Yapan Yapım Ekleri *25*
2.3.3.3. Fiilden İsim Yapan Yapım Ekleri *25*
2.3.3.4. Fiilden Sıfat Yapan Yapım Ekleri *27*
2.3.3.5. Sıfattan Sıfat Yapan Yapım Ekleri *28*
2.3.3.6. İsimden Sıfat Yapan Yapım Ekleri *29*
2.3.3.7. Zarf Yapan Yapım Ekleri *29*
2.4. İKİ DÜZEYLİ BİÇİMBİLİMSEL (MORFOLOJİK) ÇÖZÜMLEME 30
2.5. BİÇİMBİLİMSEL BELİRSİZLİK GİDERME 32
2.6. SÖZDİZİMSEL ÇÖZÜMLEME 32
3. MALZEME VE YÖNTEM 34
3.1. ÖZBEKÇENİN BİÇİMBİLİMSEL ÇÖZÜMLEMESİ 34
3.1.1. Özbekçenin İki Düzeyli Yazım Kuralları 35
3.1.2. Özbekçenin Biçim Dizgesi 48
3.1.2.1. İsim Köklü Biçim Dizgesi *50*
3.1.2.2. Fiil Köklü Biçim Dizgesi *54*
3.2. ÖZBEKÇE İÇİN BİÇİMBİLİMSEL BELİRSİZLİK GİDERME 57
3.2.1. Özbekçe için Biçimbilimsel Belirsizlik Giderme Probleminin Tanımı 58
3.2.2. Özbekçe için Biçimbilimsel Belirsizlik Giderme ve Kullanılan Yöntem 59
3.3. ÖZBEKÇENİN SÖZDİZİMSEL ÇÖZÜMLEMESİ 61
3.3.1. Bağlamdan Bağımsız Gramer 64
3.3.2. Özbekçe Sözdizimi ve Çözümlemesi 66
3.3.2.1. İsim Tamlaması *67*
3.3.2.2. Sıfat Tamlaması *68*
3.3.2.3. Sıfat-Fiil Grubu *69*
3.3.2.4. İsim-Fiil Grubu *70*
3.3.2.5. Zarf-Fiil Grubu *70*
3.3.2.6. Tekrar Grubu *71*
3.3.2.7. Edat Grubu *72*
3.3.2.8. Bağlama Grubu *73*
3.3.2.9. Unvan Grubu *74*
3.3.2.10. Özel İsim Grubu *75*
3.3.2.11. Ünlem Grubu *75*
3.3.2.12. Sayı Grubu *76*

3.3.2.12. Birleşik Fiil Grubu *77*
4. BULGULAR **79**
4.1. NUVE STUDIO İLE ÖZBEKÇENİN BİÇİMBİLİMSEL ÇÖZÜMLEMESİ 79
4.1.1. Nuve Studio ile Uygulama için Kökler Sözlüğünün Hazırlanması 80
4.1.2. Nuve Studio ile Uygulama için Ekler Sözlüğünün Hazırlanması 81
4.1.3. Nuve Studio ile Uygulama için Yazım Kurallarının Hazırlanması 82
4.1.4. Nuve Studio ile Uygulama için Biçim Dizgesi Kurallarının Hazırlanması 84
4.1.5. Nuve Studio ile Özbekçe Biçimbilimsel Çözümleme Örnekleri 86
4.1.6. Özel Durumlar, İstisnalar ve Önerilen Çözümler ile Özbekçenin Nuve Studio ile Biçimbilimsel Çözümlemesi 88
4.2. MORPHNET İLE ÖZBEKÇE İÇİN BİÇİMBİLİMSEL BELİRSİZLİK GİDERME 93
4.2.1. MorphNet ile Uygulama için Veri Setinin Hazırlanması 94
4.2.2. MorphNet ile Uygulama ve Örnekler 97
4.3. NUVE PARSER İLE ÖZBEKÇENİN SÖZDİZİMSEL ÇÖZÜMLEMESİ 99
4.3.1. Özbekçe Gramer Kurallarının Hazırlanması 103
4.3.2. Nuve Parser için Özbekçe Veri Setinin Hazırlanması 109
4.3.3. Nuve Parser için Özbekçe Doğrulama Kurallarının Hazırlanması 114
4.3.4. Nuve Parser ile Özbekçe Sözdizimsel Çözümlemesinin Çıktıları 115
4.3.5. Nuve Parser ile Özbekçe Sözdizimsel Çözümleme Örnekleri 117
4.3.6. Özel Durumlar, İstisnalar ve Önerilen Çözümler ile Özbekçenin Nuve Parser ile Sözdizimsel Çözümlemesi 123
5. TARTIŞMA VE SONUÇ **126**
KAYNAKLAR **129**
EKLER **132**
ÖZGEÇMİŞ **133**

ŞEKİL LİSTESİ

Sayfa No

Şekil 3.1: İsim kökleri için sonlu durum makinesi (FSM) modeli. 53

Şekil 3.2: Birleşik isimlerin biçim dizgesi için sonlu durum makinesi. 54

Şekil 3.3: Fiil kökleri için sonlu durum makinesi (FSM) modeli. 56

Şekil 3.4: Ahmed qora kitobni o'qidi (Ahmet siyah kitabı okudu) cümlesinin çözüm ağacı. 66

Şekil 3.5: Gullar malikasi (Çiçeklerin melikesi) – Belirtisiz isim tamlaması. 67

Şekil 3.6: Bizning uyimiz (Bizim evimiz) – Belirtili isim tamlaması. 68

Şekil 3.7: Onasini bolasi (Annesinin çocuğu) – Belirtili isim tamlaması (istisna) 68

Şekil 3.8: Yaxshi xarakterli odam (İyi karakterli insan) – Sıfat tamlaması. 69

Şekil 3.9: Xizmat ko'rsatgan kishilar (Emek sarfeden kişiler) – Sıfat-fiil grubu. 69

Şekil 3.10: Aytuv yaxshi (Söylemek iyidir) – İsim-fiil grubu. 70

Şekil 3.11: U kelganda eshitdi (O geldiğinde işitti) - Zarf-fiil grubu. 71

Şekil 3.12: Katta-katta tog'lardan keldi (Büyük büyük dağlardan geldi). 72

Şekil 3.13: Toshkentga shuning uçkun keldim (Taşkent'e şunun için geldim) – Edat grubu. 73

Şekil 3.14: Insonning tafakkuri va aqli (İnsanın tefekkürü ve aklı) – Bağlama grubu. 74

Şekil 3.15: Nolon xonim (Nalan hanım) – Unvan grubu. 75

Şekil 3.16: Muhammad Muso al-Xorazmiy (Muhammed Musâ el-Harezmi). 75

Şekil 3.3.17: Ey gardoshlar (Hey kardeşler)! – Ünlem grubu. 76

Şekil 3.18: Uch yuz qirq bir (Üç yüz kırk bir) – Sayı grubu. 77

Şekil 3.19: Odam kulib yubordi (Adam gülüverdi) – Birleşik fiil grubu. 78

Şekil 4.1: Nuve Studio için hazırlanan yazım kurallarından bir kesit. 84

Şekil 4.2: Nuve Studio için hazırlanan biçim dizgesi kurallarından bir kesit. 86

Şekil 4.3: Oyunun afişte kalması için başarıyla oynanması gerekir, cümlesinin CoNNL-u formatında gösterimi. 96

Şekil 4.4: Biçimbilimsel belirsizlik giderme için hazırlanan veri setinden bir kesit. 97

Şekil 4.5: “Men mebelsiz xotin olaman” cümlesinin uygulama çıktısı. 98

Şekil 4.6: Veri seti içerisinde yer almayan “dadam kelti (babam geldi)” cümlesinin uygulama çıktısı. 98

Şekil 4.7: Veri setinde cümle içerisinde yer alan bir kelimenin çözümü. 99

Şekil 4.8: Nuve Parser sözdizimsel çözümleyici akış diyagramı. 101

Şekil 4.9: Özbekçede yüklemi oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit. 105

Şekil 4.10: Özbekçede fiil unsurlarını oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit. 107

Şekil 4.11: Özbekçede isim unsurlarını oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit. 108

Şekil 4.12: Özbekçede soru eki unsurlarını oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit. 109

Şekil 4.13: Örnek cümlenin Nuve Parser çıktısı. 116

Şekil 4.14: Nuve Parser web tabanlı sözdizimsel ağaç üretici. 117

Şekil 4.15: Örnek olumlu isim cümlesi çözüm ağacı. 118

Şekil 4.16: Örnek olumsuz isim cümlesi çözüm ağacı. 119

Şekil 4.17: Örnek soru isim cümlesi çözüm ağacı. 119

Şekil 4.18: Örnek olumlu fiil cümlesi çözüm ağacı. 120

Şekil 4.19: Örnek olumsuz fiil cümlesi çözüm ağacı. 120

Şekil 4.20: Örnek soru fiil cümlesi çözüm ağacı. 121

Şekil 4.21: Örnek kurallı cümlenin çözüm ağacı. 121

Şekil 4.22: Örnek devrik cümlenin çözüm ağacı. 122

Şekil 4.23: Örnek basit cümlenin çözüm ağacı. 122

Şekil 4.24: Örnek birleşik cümlenin çözüm ağacı. 123

TABLO LİSTESİ

Sayfa No

Tablo 2.1: Özbek Latin alfabesi ve sembol. 11

Tablo 2.2: Özbekçe sesli harfler. 11

Tablo 2.3: Özbekçe sessiz harfler. 11

Tablo 2.4: Gramatik kategoriler ve harf kodlaması. 17

Tablo 2.5: İsim çekim eklerinin morfolojik gösterimi, kodlaması ve örneklendirmesi. 18

Tablo 2.6: Ünsüzle biten kelimelere eklenen iyelik ekleri ve örnekleri. 21

Tablo 2.7: Ünlüyle biten kelimelere eklenen iyelik ekleri ve örnekleri. 22

Tablo 2.8: İsim hâl ekleri. 23

Tablo 2.9: İsimden isim yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 24

Tablo 2.10: Sıfattan isim yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 25

Tablo 2.11: Fiilden isim yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 26

Tablo 2.12: Fiilden sıfat yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 27

Tablo 2.13: Sıfattan sıfat yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 28

Tablo 2.14: İsimden sıfat yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 29

Tablo 2.15: Zarf yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi. 30

Tablo 4.1: Nuve Studio için hazırlanan kökler sözlüğünden bir kesit. 81

Tablo 4.2: Nuve Studio için hazırlanan ekler sözlüğünden bir kesit. 82

Tablo 4.3: Tag ve eklerin eşleştirilmesini sağlayan çevirim tablosundan bir kesit. 113

SİMGE VE KISALTMA LİSTESİ

Kısaltmalar	Açıklama
CFG	: Context Free Gramer (Bağlamdan Bağımsız Gramer)
CVG	**:** Compound Verb Group (Birleşik Fiil Grubu)
FSM	: Finite State Machine (Sonlu Durum Makinesi)
FST	: Finite State Transducers (Sonlu Durum Transedürü)
GPU	: Graphics Processing Unit (Grafik İşlemci Ünitesi)
IG	**:** Interjection Group (Ünlem Grubu)
LSTM	**:** Long Short-Term Memory (Uzun-Kısa Vadeli Bellek)
JC	: Adjective Complement (Sıfat Tamlaması)
LC	: Left Context (Sol Bağlam)
NC	: Noun Complement (İsim Tamlaması)
NG	**:** Number Group (Sayı Grubu)
PNG	**:** Personal Name Group (Özel İsim Grubu)
POS	**:** Part of Speech (Sözcük Türleri)
Que	**:** Question (Soru)
RC	: Right Context (Sağ Bağlam)
RG	**:** Reduplication Group (Tekrar Grubu)
RNN	**:** Recurrent Neural Network (Yinelemeli Sinir Ağı)
S	**:** Sentence (Cümle)
SS	**:** Sub-sentence (Alt cümle)
VAG	**:** Verbal Adverb Group (Zarf-Fiil Grubu)
VJG	: Verbal Adjective Group (Sıfat-Fiil Grubu)
VNG	: Verbal Noun Group (İsim-Fiil Grubu)
VP	: Verb Phrase (Fiil Unsuru)

ÖZET

[DOKTORA TEZİ]

[Özbekçe'nin Biçimbilimsel ve Sözdizimsel Çözümlemesi ve Özbekçe-Türkçe Makine Çevirisi Sistemi]

[Ahmet Eralp KÜÇÜK]

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

[Bilgisayar Mühendisliği Anabilim Dalı]

[Danışman : Doç. Dr. Atakan KURT]

[Bu tezde, doğal dil işlemenin konularından biçimbilimsel çözümleme, biçimbilimsel belirsizlik giderme ve sözdizimsel çözümleme üzerine Özbekçe için çalışmalar yapılmıştır. Bu çalışmada önerilen yöntemlerin uygulama ve testleri çalışma kapsamında geliştirilen ve düzenlenen uygulama araçlarıyla tamamlanmıştır. Özbekçenin biçimbilimi (morfolojisi), iki düzeyli yazım kuralları ve sonlu durum makineleri ile ifade edilmiştir. Özbekçenin iki düzeyli morfolojisi bu tez için geliştirilen Nuve Studio (www.nuvestudio.com) ile test edilerek uygulaması yapılmıştır. Biçimbilimsel çözümleme sonrası ortaya çıkan belirsizlik durumlarının çözümlenmesi için yinelemeli sinir ağı modeli kullanan bir yöntem ile Özbekçenin biçimbilimsel belirsizlik giderme konusu araştırılıp, uygulaması yapılmıştır. Özbekçenin biçimbilimsel çözümlemesi ve belirsizlik giderme sonrasında elde edilen veri seti kullanılarak, sözdizimsel çözümlemesi yapılmıştır. Sözdizimsel çözümleme kapsamında Özbekçe grameri bağlamdan bağımsız gramer kuralları ile ifade edilmiştir. Cümle yapıları, kelime grupları ve gramer kuralları kullanılarak bu çalışma için geliştirilen Nuve Parser ile uygulama ve testleri yapılmıştır. Özbekçenin biçimbilimsel çözümlemesi, biçimbilimsel belirsizlik giderme ve sözdizimsel çözümlemesinde bu tez için hazırlanan 30.000 kelimeden oluşan bir veri seti kullanılarak, tekniğin bilinen durumuna göre kabul edilebilir sonuçlar sağlanmıştır.]

Şubat 2022, [146] sayfa.

Anahtar kelimeler: [Biçimbilimsel Çözümleme, Biçimbilimsel Belirsizlik Giderme, Doğal Dil İşleme, Sözdizimsel Çözümleme]

SUMMARY

Ph.D. THESIS

A Morphological and Syntactic Analysis of Uzbek and an Uzbek-Turkish Morphological Machine Translation System

Ahmet Eralp KUCUK

Istanbul University-Cerrahpasa

Institute of Graduate Studies

Department of Computer Engineering

Supervisor : Assoc. Prof. Dr. Atakan KURT

In this thesis, morphological analysis, morphological disambiguation and syntactic analysis which are the subjects of natural language processing have been studied for Uzbek. In this study the proposed methods have been completed via the application tools developed and modified within the scope of the study. Uzbek morphology is described with two level orthography rules and finite state machines. Two level morphology of Uzbek is tested and applied via Nuve Studio (www.nuvestudio.com) which is developed for this thesis. Morphological disambiguation is required after morphological analysis which produce more than one morphological analysis for a word. Morphological disambiguation of Uzbek has been investigated and applied with a method which uses recurrent neural network model. Syntactic analysis of Uzbek language is implemented by using dataset which is obtained after morphological analysis and disambiguation. Uzbek grammar is expressed with context free grammar rules. The application and tests are made with Nuve Parser which is developed for this study by using Uzbek sentence strucutres, word groups and grammatical rules. In the morphological analysis, morphological disambiguation and syntactic analysis of Uzbek, acceptable results are obtainened according to state-of-the art by using a dataset consists of 30.000 words prepared for this thesis.

February 2022, 146 pages.

Keywords: Morphological Analysis, Morphological Disambiguation, Natural Language Processing, Syntactic Analysis

1. GİRİŞ

Türki Diller ve Türkçe üzerinde Doğal Dil işleme alanında yapılan çalışmalar gün geçtikçe artmakta olup, bu çalışmaların artırılması için Türkiye'de önde gelen üniversitelerde Bilgisayar Mühendisliği Bölümlerinde bilimsel araştırmalar yapılmaktadır. Türki Dillerin alfabetik sembol birliğinin sağlanması, bu tür araştırmaların Türkiye'de ve Türki Cumhuriyetlerde yapılmasına olanak sağlamıştır. Bu tezde ele alınan Özbekçe için, Latin alfabesinin Kiril alfabesinin yerini alması ve alfabetik sembol birliğinin sağlanması ile bu tür çalışmalar hız kazanmıştır.

Çağatay Türkçesinin devamı olarak Özbek Türkçesi, Sovyetlerin kurulmasından sonra ortaya çıkan Özbekistan Cumhuriyeti'nin resmi dili haline gelmiştir. Dilin alfabesi tarih boyunca Arap, Kiril ve Latin alfabeleri olarak kullanılmış ve değişime uğramıştır. 1930'lu yıllara kadar Arap alfabesi, 1930 ve 1940 yılları arasında Latin alfabesi kullanılmıştır. 1940'lardan sonra Kiril alfabesi kullanılmaya başlanmıştır. Kiril alfabesinin kabulü yıllarında yazı dilinde kaynak olarak Farsça tesirinde kalmış olan Taşkent-Fergana vadisi ağızları esas alınması sonucunda Özbek yazı dilinde ses uyumları büyük ölçüde kaybolmuştur (Ercilasun, 2007). Çağdaş Özbek dilinde Latin alfabesi resmi alfabe olarak kabul edilmesine rağmen Kiril alfabesi de kullanılmaktadır. Bu çalışmada modern Özbekçenin 1993 yılından sonra kabul edilen Latin alfabesi (Üşenmez, 2017) referans alınmıştır.

Doğal Dil İşleme (DDİ) ya da bilgisayarlı dil bilimi, dillerin bilgisayarlar ile işlenmesini sağlayan bilim alanıdır. Dilin özelliklerinin, kurallarının ve yapısının bilgisayara tanıtılması ve farklı metodolojiler ile öğretilmesi esastır. Bu bağlamda dört ana konuda, ses bilimi, biçim bilimi, söz dizimi ve anlam bilimi konularında ilgili dil üzerinde çalışılarak dilin özellik ve kurallarını ortaya konulmaya çalışılmaktadır (Adalı, 2016). Doğal dil işlemenin ana konularından olan biçimbilimsel (morfolojik) ve sözdizimsel çözümleme bu çalışma kapsamında ele alınmıştır. Diğer Türki diller gibi sondan eklemeli bir dil olan Özbekçenin, tarih boyunca değişime uğraması ve yazı dilindeki kayıpları, biçimbilimsel ve sözdizimsel çözümleme için birçok istisnanın oluşmasına ve bu konunun daha derin araştırılması gerekliliğini ortaya koymuştur.

Bu tezde Türki Dillerin 6 ana cumhuriyet lehçelerinden (Türkçe, Türkmence, Azerice, Kırgızca, Özbekçe ve Uygurca) bir tanesi olan Özbekçenin biçimbilimi (morfolojisi), iki düzeyli yazım kuralları ve sonlu durum makineleri ile ifade edilmeye çalışılmıştır. Özbekçenin iki düzeyli morfolojisi bu tez kapsamında geliştirilen Nuve Studio[1] uygulama ortamında (Küçük, 2021) test edilerek uygulaması yapılmıştır.

Tezin bölümünde aşamasında Nuve Studio ile biçimbilimsel çözümlemesi tamamlanan ve çıktı olarak birden fazla biçimbilimsel çözümlemeye sahip kelimelerin o cümle içerisindeki en doğru çözümünün bulunması için biçimbilimsel belirsizlik giderme yapılmıştır. Morfolojik belirsizlik giderme ile morfolojik çözümlemeler ayrıştırılmıştır.

Tezin üçüncü bölümünde dilin sözdizimsel çözümlemesi yapılarak, sözdizimsel çözümleme bu tez kapsamında geliştirilen Nuve Parser ile gerçekleştirilip test edilmiştir. Özbekçenin sözdizimsel çözümlemesi için Özbekçe grameri bağlamdan bağımsız gramer(context free grammar) kuralları ile ifade edilmiştir.

Bu bölümde doğal dil işleme ve bilgisayarlı dil bilimi alanında yapılan farklı çalışmalar ile doğal dil işleme alanında kullanılan terimbilim kelimelerin açıklamalarına değinilmiştir. Tez kapsamında çalışılan problem detaylandırılmıştır. Doğal dil işlemenin ana konularından olan biçimbilimsel ve sözdizimsel çözümleme ile ilgili araştırmalara, farklı Türk dillerinde ve Özbekçe için yapılan çalışmalara yer verilmiştir.

1.1. DOĞAL DİL İŞLEME ÇALIŞMALARI

Doğal dil işleme çalışmaları, gündelik hayatımızda sıklıkla kullanılan ve hayatı kolaylaştıran bir çok bilgisayar programında, online sistemlerde ve mobil uygulamalarda yer almaktadır. Verilen bir metnin özetinin çıkarılması, konuşmayı yazıya ve yazıyı konuşmaya çeviren, metin içinde arama yapılması sağlayan, bul ve değiştir, yazım yanlışlarını düzelten, metin üzerinde analiz yapan (olumlu ve olumsuz gibi), otomatik yazışma ve diller arası çeviri uygulamaları bazı doğal dil işleme uygulamalarına örnek gösterilebilir.

Doğal dil işleme çalışmalarında dilin özellik ve kurallarının çalışılması gerekmektedir. Dil bilimciler, dilin yapısını ortaya koyan dört ana konuda, ses bilimi, biçim bilimi, söz dizimi ve

[1] https://www.nuvestudio.com

anlam bilimi konularında çalışma yaparak dilin özellik ve kurallarını ortaya koymaya çalışırlar (Adalı, 2016). Dilin özellik ve kuralları işlenerek o dil üzerinde doğal dil işleme uygulamaları yapılmaktadır. Dilin doğası gereği ortaya çıkan karmaşıklık, kuralsız durumlar, istisna ve belirsizlikler doğal dil işleme uygulamalarının karşılaştığı problemlerin başında gelmektedir.

Özbekçe üzerinde yapılan Doğal Dil İşleme çalışmaları gün geçtikçe artmakta olup tezin ilgili çalışmalar bölümünde anlatılmaktadır. Bu tezde sunulan ve Özbekçenin biçimbilimsel ve sözdizimsel çözümlemesi problemine ve ele alınan çözümlere terminoloji bölümden sonra değinilmiştir.

1.2. TERMİNOLOJİ

Doğal Dil İşleme alanında kullanılan terimbilim kelimelerin açıklamaları ve tezde kullanılan tanımlamaları aşağıdaki gibidir.

Morfem ya da biçimbirim, dildeki anlam ayırıcı özelliğe sahip en küçük dilsel birim (İmer ve diğ., 2011), çoğu ek halinde olan kelime parçalarıdır.

Morfoloji ya da biçimbilim, dilbiliminde sözcüklerin içyapısını inceleyen alt dalıdır (İmer ve diğ., 2011). Temel inceleme nesnesi, dilin anlam taşıyan en küçük parçaları olan biçimbirimlerdir. Biçimbilim ya da morfoloji, sözcükleri, nasıl oluşturulduklarını ve diğer sözcüklerle ilişkilerini inceler ve sözcüklerin kök, gövde ve ek gibi bileşenlerinin yapısını çözümler.

Morfotaktik ya da biçim dizgesi, morfemlerin sıralanmasına ilişkin düzen kısıtlamalarını temsil eder. Morfemlerin birbirini nasıl takip edeceklerini, sıralamalarını ve düzenini tanımlayan kurallar dizisidir.

Ortografi, yazım ile ilgilidir. Ortografi kuralları, yazım kuralları anlamına gelmektedir. Ortografi, yazım denetimi, heceleme, kelime sonu, kelime ayırma, vurgu ve noktalama işaretlerini içerir.

Sözcüksel (lexical) düzeyi, bir sözcüğün işlevsel bileşenlerinin yapısını belirtir ve gösterir.

Yüzeysel (surface) düzeyi, sözcük diziminin sözcük yapısının standart yazım kuralları ile gerçekleştirilmesini göstermektedir.

İki düzeyli kurallar, kelime yapısının sözcüksel (lexical) ve yüzeysel (surface) düzeyler ile gösterimine dayanmaktadır.

Biçimbilimsel çözümleme, cümle içerisindeki bir kelimenin kök ve eklerine ayrıştırılarak her bir bileşenin (morfemin) sırasıyla tanımlanması demektir.

Biçimbilimsel belirsizlik giderme, bir kelimenin, bir cümle içerisinde ortaya çıkan birden fazla biçimbilimsel (morfolojik) çözümünden, belirsizliği ortadan kaldırarak, o cümle için doğru olan, istenilen ve tek bir çıktının üretilmesini sağlayan yöntemlerdir.

Sözdizimi, bir cümlenin içerisindeki kelimelerin veya kelime gruplarının dilbilgisi kuralları doğrultusunda cümlenin içerisinde nasıl sıralanacağını tanımlar.

Sözdizimsel çözümleme, bir cümleyi oluşturan kelime ve kelime gruplarını ayrıştırarak dilbilgisi kuralları doğrultusunda cümlenin içerisinde nasıl sıralandığını göstermektedir. Cümle içerisindeki kelime ve kelime gruplarının birbirleriyle olan ilişkini tanımlamaktadır.

Bağlamdan bağımsız gramer (CFG), doğal dil işlemede dil tasarımı ve dilin kurallarını tanımlamak için kullanılan biçimsel ve matematiksel bir gramer türüdür.

1.3. PROBLEM TANIMI VE ÖNERİLEN ÇÖZÜM

Özbekçe üzerinde doğal dil alanında yapılan çalışmalar temel seviyede olup bu çalışmaların daha ileri bir seviyeye taşınması için son yıllarda bu alanda birçok farklı çalışma yapılmaktadır. Özbek alfabesi ve dili tarih boyunca birçok değişime uğramıştır. Bu değişimler dil ve yazı üzerinde bazı kayıplara neden olmuştur. Bu kayıplar ve değişimler doğal dil işleme çalışmalarını da etkilemiştir. Doğal dil işleme alandaki çalışmaların temel seviyede olmasında bu değişim ve kayıpların payı büyüktür.

Özbek alfabesi Arap, Latin ve Kiril alfabesi ile kullanılmış olup, günümüzde Kiril ve resmi olarak Latin alfabesi ile kullanılmaktadır. Resmi olarak Latin alfabesi ile kullanılması bu tezin Latin Özbek alfabesini referans alarak hazırlanmasına neden olmuştur. Özbekçe, diğer Türk dilleri gibi sondan eklemeli ve biçimbilimsel (morfolojik) açıdan zengin bir dildir. Özbekçede dilin ve alfabenin tarih boyunca değişimi esnasında yazı dilinde ses uyumlarının büyük ölçüde kaybolması ve düzenli yapının yanı sıra istisnaların fazla olması doğal dil alanında çalışmaları bir nebze zorlaştırmıştır. Biçimbilimsel ve sözdizimsel çözümlemenin yapılabilmesi için

tanımlanan kuralların ve istisnaların detaylı bir şekilde çalışılması gerekmektedir. Bu çalışmanın her bölümünde problemin daha alt katmanlara ayrılarak çözümünün sağlanması yöntemleri kullanılmıştır. Bu doğrultuda sözdizimsel çözümleme için biçimbilimsel çözümlemeye ve biçimbilimsel belirsizlik gidermeye ihtiyaç duyulmaktadır. Tezim ilk kısmında Özbekçenin biçimbilimsel çözümlenmesi ve bulguları tamamlanarak, ikinci kısmında biçimbilimsel belirsizlik giderme yapılarak son kısmında sözdizimsel çözümlemeler üzerinde çalışmalar başlatılmıştır.

Özbekçenin biçimbilimsel çözümlenmesi için, Özbekçenin biçimbilimi (morfolojisi) 26 iki düzeyli yazım (ortografik) kuralları ile ifade edilmiştir. Özbekçe kelime yapılarının biçim dizgeleri (morfotaktikleri) fiil ve isim paradigmaları için sonlu durum makinaları ile gösterilmiştir. Bu çalışmada Özbekçenin iki düzeyli biçimbilimsel çözümlemesinin uygulamasında ilk olarak Nuve (Zafer, 2018) biçimbilimsel çözümleyicisi kullanılmıştır. Nuve çözümleyicisinin kullandığı, Nuve açık kaynak kodlu biçimbilimsel analiz ve makine çevirisi çatısına ait kütüphaneler geliştirilerek, ihtiyaçlara uygun olarak www.nuvestudio.com URL adresinden erişilebilen Nuve Studio çalışmanın sonraki bölümlerinde kullanılmıştır. Uygulamanın Özbekçe yazım kurallarında ve istisnalarında daha doğru çalışması için bu geliştirme ihtiyaç olarak görülmüş ve yeni bir sistem bu çalışma için geliştirilmiştir. Uygulama aşamasında Özbekçe için 30.000 kelime kök sözlüğü olarak ve 370 ek ise ek sözlüğü olarak kullanılmıştır. Özbekçenin biçimbilimsel çözümlenmesi probleminde elde edilen bulgular ile bir sonraki aşamaya geçilmiş ve biçimbilimsel belirsizlik giderme ihtiyacı ortaya çıkmıştır. Biçimbilimsel çözümlemelerden sonra birden fazla çözümün üretilmesi ve istenilen (en doğru) çözümün gösterilmesi için belirsizlik giderme aşamasına geçilmiştir.

Biçimbilimsel çözümleme için verilen cümlelerin içerisinde yer alan kelimelerin birden fazla çözüme sahip olması biçimbilimsel belirsizlik problemini ortaya çıkarmıştır. Biçimbilimsel belirsizlik problemi için biçimbilimsel belirsizlik giderme işlemleri uygulanmıştır. Biçimbilimsel belirsizlik giderme; bir kelimenin (bir metindeki veya yalın), birden fazla biçimbilimsel çözümlemesi olması durumunda ilgili kelimenin biçimbilimsel çözümünün tek bir çözüme, istenilen(doğru) çözüme indirgenmesi ve tek bir çözümün çıktı olarak gösterilmesini sağlayan yöntemlerin adıdır. Dayanık ve diğ. (2018) önerdikleri MorphNet modeli ile girdi olarak aldığı metnin içerisindeki kelimelerin doğru biçimbilimsel çözümlemesini çıktı olarak üreten, bir sıralı (seq2seq) tekrarlayan yapay sinir ağı (RNN) modeli

sunmuşlardır. Bu model, Özbekçenin biçimbilimsel belirsizlik giderilmesi için kullanılmış olup, öncelikli olarak 13.000 kelimelik bir veri seti ile sonra 30.000 kelimelik bir veri seti ile uygulama gerçekleştirilmiştir. Biçimbilimsel belirsizlik giderme işleminden sonra sözdizimsel çözümleme aşamasına geçilmiştir.

Tezin son bölümünde dilin sözdizimsel çözümlenmesi problemi üzerine araştırmalar, çalışmalar ve geliştirmeler yapılmıştır. Özbekçenin sözdizimsel çözümlemesi için "Yukarıdan – Aşağıya (Top – Down)" sözdizimsel çözümleyici yaklaşımına sahip, morfolojik çözümlemeye, bağlamdan bağımsız gramer kurallarına ve geçerlilik-doğrulama kurallarına dayanan bir çalışma referans (Zafer, 2011) alınarak bu çalışmadaki yöntem ve çözümleyici kullanılmıştır. Özbekçe cümle türleri ve kelime grupları çalışılarak uygulama aşamasında 286 gramer kuralı tanımlanmıştır. Referans çalışmada açık kaynak kodlu olarak sunulan sözdizimsel çözümleyici geliştirilmiştir. Nuve Parser ismini verdiğimiz çözümleyici Nuve projesinin sözdizimsel çözümleyicisidir. Uygulama aşamasında 17.000 kelimeden oluşan bir veri seti kullanılmış olup, Nuve Parser'a tanımlanan bir cümle ögelerine ayrıştırılarak tanımlanan gramer kurallarına uygun olarak çıktı üretmektedir.

Doğal dillerin yapısı gereği ortaya çıkan ve bilgisayarlı dil biliminde ifade edilmeyi zorlaştıran kuralsız durumlar, istisnalar ve belirsizlikler çözümlenmesi gereken problemlerdir. Bu çalışma ile Özbekçenin biçimbilimsel ve sözdizimsel çözümlemesi problemine, detayları tezin "Malzeme ve Yöntem" bölümünde anlatılan çözümler sunulmuş, uygulamaları geliştirilen yazılımlar ile yapılmış ve test sonuçları ilgili bölümde paylaşılmıştır.

1.4. İLGİLİ ÇALIŞMALAR

Bu bölümde tezin "Doğal Dil İşleme Çalışmaları" bölümünde anlatılan uygulamalara ve örnek yöntemlere değinilmiş, bu tezde öne sürülen çalışmalar ile ilgili çalışmalara değinilmiş, Türki dillerde ve Özbekçede bu tez kapsamında yapılan bazı referans çalışmalar sunulmuştur.

Günlük hayatta kullanılan doğal dil işleme uygulamalarından birisi, verilen bir metinin özetinin çıkarılmasıdır. Tülek (2007) farklı istatistiksel yöntemler ile Türkçe için metin özetlemesi gerçekleştirmiştir. Konuşmayı yazıya çeviren sistemler birçok mobil ve mesajlaşma uygulamasında kullanılarak pratiklik sağlamaktadır. Yazıyı konuşmaya çeviren uygulamalar bir metinin sese dönüştürülmesini sağlamaktadır. Yazıyı konuşmaya ve konuşmayı yazıya çeviren farklı teknik ve algoritmalar gözlemlenerek bulguları sunulmuştur (Trivedi ve diğ.,

2018). Metin içerisinde arama yapılmasının sağlanması araştırmaları kolaylaştırmakta ve bugün hali hazırda kullanılan birçok web arama motorunun temelini oluşturmaktadır. Metin içerisindeki bir bölümün, cümlenin ve kelimenin aranması gibi konular doğal dil işlemede bilgi çıkarımının konusudur. Lewis ve Jones (1996), metin içerisindeki arama performansının artırılması için mevcut çözümleri incelenerek istatistiksel bir çözüm sunmuşlardır. Metin içerisinde aranan kelimenin bulunduktan sonra bir başka kelime ile değiştirilmesini sağlayan bul ve değiştir uygulamaları ve onların performansları ayrıca araştırılmıştır. Bir dile ait özel çalışmalar performansı daha da artıracağı öngörüsü ile farklı çalışmalar yapılmıştır. Şentürk ve Adalı (2016), Türkçe için kural tabanlı bir bul ve değiştir algoritması geliştirmiş ve uygulamasını yapmıştır.

Bu tezde araştırması yapılan, biçimbilimsel (morfolojik) ve sözdizimsel çözümleme doğal dil işlemenin ana konuları arasında yer almaktadır. Tezin birinci bölümünde, Özbekçenin morfolojisi üzerine açıklamalara yer verilmiş olup, iki düzeyli Özbek Morfolojisi detaylandırılmıştır.

Özbekçenin iki düzeyli morfolojisinin tanımı ve analizi için Koskenniemi (1983)'nin iki düzeyli morfoloji modeli referans alınmıştır. Birçok sondan eklemeli dilin biçimbilimsel tanımlamasında Koskenniemi (1984)'nin dilden bağımsız kelimenin üretimi ve tanımlanması sağlayan iki düzeyli modeli kullanılarak yazım kuralları iki düzeyli kurallarla tanımlanır ve biçim dizgesi (morfotaktik) kuralları sonlu durum makineleri ile gösterilir. Özbekçe üzerinde hazırlamış olduğumuz çalışmanın temelini Koskenniemi'nin bu çalışmaları oluşturmaktadır. Ayrıca birçok çalışmada farklı uygulama araçları kullanılarak bu model uygulanmıştır. Bask ve Svahili dilleri karmaşık morfolojiye sahip sondan eklemeli dillerdir. Svahili dilinin morfolojik analizi için iki düzeyli morfoloji modeli kullanılmış olup, geliştirilen SWATWOL uygulama aracı ile testleri yapılmıştır (Hurskainen, 1992). Bask dilinin biçimbilimsel çözümlemesi için iki düzeyli morfoloji modeli ile XUXEN uygulama aracı kullanılmıştır (Agirre ve diğ., 1992). Japonca ve Korece de sondan eklemeli karmaşık morfolojiye sahip dillerdir. Japoncanın morfolojik analizi de Koskenniemi'nin iki düzeyli morfolojik modeli ve morfolojik çözümleyici programı KIMMO kullanılarak yapılmıştır (Alam, 1983). İki düzeyli Korece morfolojisi KIMMO programının gelişmiş versiyonu olan PC-KIMMO (Antworth, 1990) morfolojik çözümleyici ve üretici programı kullanarak uygulanmış ve test edilmiştir (Kim ve diğ., 1994).

Türk dilleri de yukarıda bahsedilen diller gibi sondan eklemeli ve karmaşık bir morfolojiye sahip dillerdir. İki düzeyli morfoloji modeli Türkçe ve Türk dillerinin biçimbilimsel çözümlemesinde kullanılmıştır. Türkçe morfolojisi için bu alanda yapılan öncü çalışmalardan birinde, Oflazer (1994) iki düzeyli yaklaşımla, 22 iki düzeyli yazım kurallarını tanımlayarak, PC-KIMMO morfolojik çözümleyici ve üretici programını kullanarak Türkçenin iki düzeyli morfolojik tanımını yapmıştır. Koskenniemi'nin iki düzeyli modeli ile morfolojik çözümlemenin gerçekleştirildiği birçok Türki dil bulunmaktadır. Türk dillerinden Kazakça, Tatarca ve Kırgızca için bu yaklaşım ile gerçekleştirilen bazı örnek çalışmalar sırasıyla şöyledir. Zafer ve diğ. (2011), Kazak morfolojisinin iki düzeyli tanımı üzerine çalışmalar yapmış ve Dilmaç çatısı (framework) üzerinde uygulama ve testlerini tamamlamıştır. Gökgöz ve diğ. (2015), Tatarcanın iki düzeyli morfolojisi üzerinde çalışmış, Tatarca biçim dizgesi için dil (ortografi) kurallarını tanımlayarak, sonlu durum makineleri ile gösterimini sağlamışlardır. Tatarca uygulaması için Nuve çatısı (framework) kullanılmıştır. Kırgız morfolojisi, iki düzeyli model yaklaşımı ile tanımlanmış ve Nuve çatısı (framework) kullanılarak uygulaması yapılmıştır (Yiner ve diğ., 2016).

Özbekçede diğer Türki diller gibi sondan eklemeli ve karmaşık morfolojiye sahip bir dildir. Özbek alfabesi üzerinde tarih boyunca gerçekleşen değişimler, bu değişimler neticesinde gramerde oluşan düzenli olmayan yapılar ve Özbekçe yazı dilimde ses uyumu kayıpları gibi önemli gramer kayıpları, Özbek morfolojisi üzerinde yapılan çalışmaları etkilemiştir. Modern Özbek dilinin morfolojisi ve morfolojik çözümlemesi üzerine yapılan çalışmalar sınırlı sayıda olup, bu çalışmalardan bazıları Özbekçe için doğal dil işleme konusunda önemli bir yere sahiptir. Bir dilin biçimbilimsel çözümlemesinin yapılabilmesi için o dilin yazım kurallarına ihtiyaç vardır. İstisnalar azalması ve düzenli yazım kurallarının olması o dilde biçimbilimsel çözümlemenin başarısını artırmaktadır. Bu kadar fazla değişimden sonra Özbekçenin yazım kuralları için kilometre taşı sayılabilecek bir karar alınmıştır. Çağdaş Özbekçenin yazım kuralları 24 Ağustos 1995 tarihinde Özbekistan Bakanlar Kurulu'nun 339 sayılı kararı ile yayınlanmıştır (The Uzbekistan Cabinet of Minister's Resolution 1995, No.339). Bu günümüz Özbekçesi için son derece önemli bir çalışmadır. Modern Özbekçe için tanımlanan bu kurallar dil çalışmalarında ana referans kabul edilmiştir. Matlatipov ve Vetulani (2009), tarafından Özbek morfolojisi, biçim dizgesi (morfotaktik) ve morfofonemik kurallarla sunulmuştur. Tanımlanan Özbek morfolojisinin çözümlenmesi ve analizi için prolog tabanlı UZMORPP sistemi kullanılmış ve uygulaması yapılmıştır. Atadjanov (2016), Özbekçe kelimelerin

biçimbilimsel analizini FST metot ve yaklaşımıyla sağlayarak kök kelime tanımı ve kelime sıralaması gösterilmiştir. Abdurakhmonova (2019), Özbekçenin iki düzeyli biçimbilimsel analizi FST teknoloji tabanlı biçimbilimsel çözümleyici ile uygulamıştır.

2. GENEL KISIMLAR

Tezin bu bölümünde biçimbilimsel ve sözdizimsel çözümlemesi yapılan Özbekçe'nin ses kuralları, dil hakkında genel bilgiler, biçimbilimsel ve sözdizimsel çözümleme ile alakalı genel bilgilendirmeler sağlanmıştır.

2.1. ÖZBEK DİLİNE GENEL BAKIŞ VE ÖZBEKÇE GENEL BİLGİLER

Özbek halkı, günümüzde Sovyetlerin kurulmasıyla ortaya çıkan Özbekistan Cumhuriyetinde yaşamaktadır. Özbek halkının tarihi, 14. Yüzyıla kadar dayanmaktadır. Altınordu devletinin prenslerinden biri olan Özbek Han'ın Kazak bozkırından Türkistan'a doğru ilerlemesiyle günümüz Özbekistan coğrafyasına gelmişlerdir. Özbek Hanı aslen bir Kıpçak Hanıdır. Sovyetler kurulduktan sonra da bu adın kullanılmasıyla beraber Özbekistan ismi bir Cumhuriyetin adı olmuş, Özbekistan Cumhuriyetinde yaşayan halklara Özbekler ve dillerine Özbekçe ya da Özbek Türkçesi denilmeye başlanmıştır.

Çağatay Türkçesinin (Türkiy Til, Türkî Til) devamı olarak Özbek Türkçesi, Sovyetlerin kurulmasından sonra ortaya çıkan Özbekistan Cumhuriyetinin resmi dili hâline gelmiştir. Bu döneme kadar alfabe olarak Arap alfabesinin kullanıldığı bilinmektedir. Sovyetler kurulduktan sonra Özbekler de bu alfabeyi 1930'a kadar kullanmışlardır. Ardından 1939'a kadar (1930-1940 yılları arasında) Özbekistan'da Latin asıllı Özbek alfabesi kullanılmıştır (Yusupova, 2018). 1939'dan sonra da Özbek Türkleri için Kiril asıllı Rus alfabesine ek olarak ў "o" ve "ö", қ *"ķ/q"*, ғ "ġ", ҳ "h" harfleri eklenerek Kiril asıllı Özbek alfabesi kabul edilmiştir. 1993'den sonra kabul edilen Özbek Latin alfabesinde, 29 harf ve apostrof karakteri ile 30 karakter bulunmaktadır (Üşenmez, 2017). Özbek Latin alfabesi Tablo 2.1'de gösterilmektedir. Latin alfabesine geçişin kabulü ile tez çalışmamızda Özbek Türkçesi için Latin alfabesi tercih edilmiştir. Bu nedenle Özbekçenin biçimbilimi Latin alfabesi kullanılarak ifade edilecektir.

Bu alfabede Özbekçe ünlü sesleri (seslileri – ünlüleri) karşılayan 6 harf vardır. Bu ünlüler sırasıyla, "a", "e", "i", "u", "o" ve " o' " harfleridir. Ünsüz sesleri (sessizleri – ünsüzleri) karşılayan 23 harf bulunmaktadır. Bu ünsüzler sırasıyla, "b", "d", "f", "g", "h", "j", "k", "l", "m", "n", "p", "q", "r", "s", "t", "v", "x", "y", "z", " g' ","sh", "ch","ng" harfleridir. Özbek Latin alfabesi 29 harf ve apostrof (') karakteri ile 30 karakterden oluşmaktadır. Özbekçe

alfabesi Tablo 2.1'de, ünlü harfleri Tablo 2.2'de ve Özbekçe ünsüz harfleri Tablo 2.3'de fonetik özellikleriyle gösterilmiştir.

Tablo 2.1: Özbek Latin alfabesi ve sembol.

A a	G g	L l	Q q	V v	G' g'
B b	H h	M m	R r	X x	Sh sh
D d	I i	N n	S s	Y y	Ch ch
E e	J j	O o	T t	Z z	Ng ng
F f	K k	P p	U u	O' o'	'

Tablo 2.2: Özbekçe sesli harfler.

	Yuvarlak			**Düz**		
Ünlüler	**Açık**	**Yarı Kapalı**	**Kapalı**	**Açık**	**Yarı Kapalı**	**Kapalı**
Arka Sıralı	o	o'	u			
Ön Sıralı				a	e	i

Tablo 2.3: Özbekçe sessiz harfler.

		Dudaksıllar	**Dişsel**	**Damaksıl**	**Art Damak**	**Küçük Dil**	**Gırtlaksal**
Süreksiz	**Tonsuz**	p	t		k	q	
	Tonlu	b	d		g		
Yarı Patlayıcı	**Tonsuz**			ch			
	Tonlu			j			
Sızıcı	**Tonsuz**	f	s	sh	x		h
	Tonlu		z			g'	
Geniz		m	n		ng		
Akıcı		v	l,r	y			

Özbekçe için iki düzeyli ortografik (yazım) kuralları tanımlayabilmek için Özbekçe dilbilgisine, ek ve kök-gövde arasındaki ses uyumlarına uygun olarak belirli harf kümeleri oluşturulması gerekir. Bu harf kümeleri ya da diğer bir deyişle meta harfler büyük harfler ile yazılarak gösterilir ve o kümeye ait bütün harfler için genel kuralların tanımlanmasını kolaylaştırmaktadır. Özbekçe yeni Latin alfabesine uygun yazım kurallarını tanımlarken kullanılacak harf kümeleri aşağıdaki gibidir.

Sessiz harfler kümesi (C); C= {b, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, z, y, x, g', sh, ch, ng},

Sert sessiz harfler kümesi (C_s); C_s= {f, s, t, k, q, ch, sh, l, n, s, h, y},

Yumuşak sessiz harfler kümesi (C_y); C_y= {b, d, g, h, j, m, p, r, v, z, x, g', ng},

Sesli harfler kümesi (V); V= {a, e, i, u, o, o'},

Ön sıralı düz sesli harfler kümesi (V_{fu}); V_{fu}={a, e, i},

Arka sıralı yuvarlak sesli harfler kümesi (V_{bo}); V_{bo}= {o, u},

Diğer harf kümeleri:

A= {a, o},

V_{de}= {o, i, u},

D= {d, t},

Q= {q, k},

G={q, k, g, g'},

I={u, i},

Y={a, y},

E={a, i}.

2.2. ÖZBEKÇE SES BİLGİSİ, SES OLAYLARI VE SES DEĞİŞİMLERİ

Bu bölüm ve alt bölümlerde Özbekçe ses olayları, kuralları ve ses değişimleri detaylandırılmıştır.

2.2.1. Hece ve Hece Türleri

Özbekçe hece ve hece türleri kelimenin biçimbilimsel çözümlenmesi ve dolayısı ile sözdizimsel çözümlemesi için incelenmesi gereken ses bilgisi konularının başında gelmektedir. Özbekçe için hece ve hece türleri aşağıdaki gibidir:

- Tek ünlü, e-şit,
- Açık hece, sonu ünlü ile biten hecedir.
- Kapalı hece, sonu ünsüz ile biten hecedir.
- Ünlü + ünsüz,
- Ünsüz + ünlü,
- Ünlü + ünsüz + ünsüz,
- Ünsüz + ünlü + ünsüz,
- Ünsüz + ünsüz + ünlü,
- Ünsüz + ünlü + ünsüz + ünsüz,
- Ünsüz + ünsüz + ünsüz + ünlü,
- Ünsüz + ünsüz + ünsüz + ünlü + ünsüz,
- Ünsüz + ünsüz + ünlü + ünsüz + ünsüz,

2.2.2. Vurgu

Özbekçede vurgu her zaman ünlünün üzerindedir. Özbekçede vurgu genel olarak kelimenin son hecesindedir. Örneğin; kitób, maktáb, qorá, saríq, yigirmá, to‘qsón vb.. Kelimelere ekler eklendikçe, vurgu da sözün son hecesine kayar. Örneğin; bolá, bolalár, bolalargá vb..

Özbek dilinde vurgu almayan bazı ögeler ve ekler vardır. Özellikle, yapım eklerinden zarf yapım ekleri -cha, -day (-dek), çekim ekleri -dir, -man, -san, -miz, -siz, omulsuzluk eki -ma ve -ku, -u, -yu, -da, -mi, -çi sona gelen edatlar vurgu almazlar (Аскарова, 1972).

2.2.3. Ses Uyumları

Özbek alfabesi tarih boyunca değişime uğramıştır ve bu değişim sırasında bazı kayıplara da uğramıştır. Kiril alfabesinin kabulü yıllarında yazı dilinde kaynak olarak Farsça tesirinde kalmış olan Taşkent-Fergana vadisi ağızları esas alınması sonucunda Özbek yazı dilinde ses uyumları büyük ölçüde kaybolmuştur (Ercilasun, 2007).

Ünlü ve ünsüz uyumlarının tutarsız olması birçok ekin tek şekilli olmasına ve ortografik kuralları oluştururken birçok istisna ile karşılaşmamıza neden olmuştur. Çekim ve yapım eklerinin kök ve gövdeye eklenmesi durumunda gerçekleşen ses uyumları ilgili alt bölümlerde detaylandırılmış, büyük ünlü uyumu, küçük ünlü uyumu ve ünsüz uyumları istisnalara ve tutarsızlığa rağmen gruplandırılarak detaylı olarak incelenmiştir.

2.2.3.1. Büyük Ünlü Uyumu

Özbekçede düzenli bir büyük ünlü uyumu aranmaz. Karluk grubu Türkçelerinde görülen temel ses bilgisi özelliklerinden bir olan "i'leşme" bu uyumu bozan esas unsurdur. Kalın sıradan ünsüzlerin yanından başka diğer yerlerde i ünlüsü yoktur, bosh-i-miz (başımız) ve qil-di-lar (kıldılar) örneklerinde bu durum gösterilmektedir.

Özbekçede büyük ünlü aranmadığını gösteren diğer bir örnek, e ve a harflerinin ayrımının net olarak yapılmamasından kaynaklanmaktadır. Bu harflerinin ayrımının net çizgilerle yapılmaması da Özbekçede büyük ünlü uyumunu bozar. Özellikle yuvarlak a'nın kullanıldığı örneklerde durum daha açıktır. Örneğin bil-ma-moq (bilmemek) ve bosh-la-moqda-miz (başlamaktayız) örneklerinde gösterildiği üzere Özbekçede büyük ünlü uyumu aranmaz.

2.2.3.2. Küçük Ünlü Uyumu

Özbekçede genel olarak ses uyumu olmadığı görülmektedir (Шоабдураҳмонов, 1980). Buna rağmen küçük ünlü uyumunun korunduğu durumlara rastlanmaktadır. Küçük ünlü uyumunun korunduğu ve küçük ünlü uyumunun bozulduğu durumlar aşağıda detaylı olarak anlatılmıştır. Küçük ünlü uyumunun bozulduğu haller aşağıdaki gibidir:

- İnce-dar-düz ünlünün yardımcı ünlü olarak tercih edilmesinden dolayı bozulur. Örneğin; kok-i-sh (mavimsi), qol-i-m (elim), yol-i-q- (rastlamak), yul-i-n- (yolunmak), uch-i-r- (uçurmak) ve uch-i-sh- (uçuşmak) örneklerinde küçük ünlü uyumu bozulmuştur, aranmaz.
- Düzlük yuvarlaklık uyumunu bozan ikinci durum, bazı eklerin özellikle dar-düz ünlülü şekillerinin kullanılmasıdır. Örneğin; kuch-li (güçlü), o't-mish (geçmiş), o't-sin (geçsin), o'l-gi-m (ölesim) ve bu-ning (bunun) örneklerinde küçük ünlü uyumu bozulmuştur, aranmaz.

Özbekçede ünlü ve ünsüz uyumlarının tam ayrılamamasına rağmen bazı durumlarda küçük ünlü uyumunun korunduğu gözlemlenmektedir. Küçük ünlü uyumunun korunduğu haller aşağıdaki gibidir:

- Ünsüz ile biten bazı fiillere -uv eki eklenince, ol-uv (çıkarma), yoz-uv (yazı) ve o'l-dir (öldürmek) örneklerinde olduğu gibi küçük ünlü uyumu korunmaktadır.

- Ünsüz ile biten ve U ünlüsü barındıran bir fiile -uq eki eklenince küçük ünlü uyumunun korunduğu gözlemlenir. Örneğin; uz-uq (kopuk) ve yut-uq (kazanç) örneklerinde olduğu gibi küçük ünlü uyumu korunmaktadır.

2.2.3.3. Ünsüz Uyumu

Özbek Türkçesinde ünlü ve ünsüz uyumları kesin çizgilerle birbirinden ayrılmamıştır. Bunun sebepleri arasında, tarih boyunca gerçekleşen alfabe değişimleri ve yoğun kültürel etkileşimler gibi faktörler sıralanabilir.

Ünsüz uyumu özellikle eklerin sedalı şekillerinin kullanılmasından dolayı bozulur. Yaygın olarak daha çok çekim ekli şekillerde, bazen de yapım ekli şekillerde bu uyumsuzluk görülür. Örneğin; Özbek-dir (Özbektir), tut-di (tuttu), mening-cha (bana göre) ve suhbat-dosh (muhabbet arkadaşı) örneklerinde olduğu gibi ek yapısında değişme gözükmediği için ünsüz uyumu bozulmuştur. Ayrıca, alıntı kelimelerde ünsüz benzeştirilmesinde ise kelimeyi aldığı dildeki ses özelliklerine göre yazılma tercih edilmiştir. Örneğin; dastgoh (tezgah), taqdim (sunma) ve kitobxona (kütüphane) kelimelerinde bu durum gösterilmektedir.

2.3. ÖZBEKÇE YAPIM VE ÇEKİM EKLERİ

Özbek Türkçesi, diğer Türki diller gibi sondan eklemeli bir dildir. Arapça ve Farsçadan dile giren baştan eklemeli kelimeler bu özelliğini kaybetmiştir. Özbekçede ekler kelimeye sondan eklenen yapım ve çekim ekleri olarak ikiye ayrılmaktadır.

İsimlerin ve fiillerin, kök veya gövdelerine eklenerek isim veya fiil türeten, anlam değişikliği sağlayabilen eklere yapım ekleri denir, yapım ekleri türetme ekleridir. İsimlerin ve fiillerin, kök veya gövdelerine eklenerek anlam değişikliği sağlamaksızın eklendiği kelimenin cümle içerisinde zaman, yön, aitlik ve çoğulluk gibi anlamlar almasını sağlayan eklere çekim ekleri denir.

Yapım ve çekim eklerinin biçimbilimsel gösterimi için farklı yöntemler kullanılmıştır. Birçok benzer çalışmada birbirinden tamamen farklı olarak eklerin biçimbilimsel gösterimi ile karşılaşılmıştır. Özbekçe, Türkçe grubuna ait diller ile benzer yapı taşıdığı için eklerin biçimbilimsel gösterimi konusunda diğer Türki diller ile ortak bir gösterime sahip olması için birçok çalışmada referans alınan Oflazer (1994) çalışmasındaki ek yapısı referans alınmıştır.

Buna göre, eklerin geldiği kelime kök ve gövdelerine göre çeşitli varyasyonları bulunmaktadır. Yani ekteki harfler eklendikleri kök ve gövdeye göre değişim gösterebilir. Olası karakterler ve bu farklılaşma meta harfleri ile gösterilmiştir. Özbekçede birinci tekil iyelik eki -im veya -um olabilir. Bu ekler -Im olarak gösterildiğinde, I meta harfi i, u ünlülerinden birine tekabül etmektedir. Bu harf ekin geldiği kelimeye göre özelleşebilir. I meta harfinden sonra gelen m harfi aynen kalır.

Bu bölümde Özbekçe yapım ve çekim eklerine değinilmiştir. Özbekçede isim gövdelerine çoğul, iyelik, hal ve yapım ekleri, fiil gövdelerine olumsuzluk, zaman, şahıs ve yapım ekleri eklenir. Bu bölümün alt bölümlerinde eklerin biçimbilimsel gösterimi için kullanılan kodlama şemasına, kodlama şeması ile gösterilen Özbekçe isim çekim, isim yapım, fiil çekim ve fiil yapım eklerine değinilmiştir.

2.3.1. Kodlama Şeması

Eklerin biçimbilimsel (morfolojik) gösterimi ve sıralanmasının açıklaması bu bölümde detaylandırılmış olup, sonraki bölümlerde kullanımı örneklerle gösterilmiştir. Buna göre, büyük harfler meta sesleri göstermektedir. Bir ekin morfolojik gösterimi P pozisyon olmak üzere açıklamaları aşağıdaki gibi ve sıralaması şu şekildedir; $P_0P_1P_2_P_3$...

P_0: Pozisyon 0'da bulunan son (final) gramatik türü ifade etmektedir. Ek kök veya gövdeye geldikten sonra sözcüğün dönüştüğü türdür. Bir fiile, fiilden isim yapan ek geldiği zaman fiil isme dönüşür ve P_0 isim olarak gösterilir.

P_1: Pozisyon 1'de ek sözcüğe eklenmeden önceki türü yani kaynak gramatik türü ifade etmektedir. Ek kök veya gövdeye eklenmeden önceki sözcüğün asıl (orijinal) türünü gösterir. Bir fiile, fiilden isim yapan ek geldiği zaman fiil isme dönüşür ve P_1 fiil olarak gösterilir.

P_2: Pozisyon 2'de sözcüğe eklenen türün yapım veya çekim eki olduğu gösterilir. I (Inflectional) çekim ekini, D (Derivational) ise yapım ekini ifade etmektedir. Bir fiile, fiilden isim yapan ek geldiği zaman fiil isme dönüşür-türetilir ve P_2, eklenen yapım eki D harfi ile gösterilir.

P_3: Pozisyon 3'de ekin lexical (sözlüksel) gösterimi ifade edilmektedir. Bir fiile, fiilden isim yapan –in, un eki geldiği zaman bu ek i ve u harflerini gösteren meta harfi I ile (I)n şeklinde gösterilir. I(n) ise P_3 'te gösterilir.

Gramatik kategoriler (türler) ve harf kodlaması Tablo 2.4 gösterilmiştir. Yukarıda tanımlanan pozisyonlara ve Tablo 2.4 gösterilen gramatik kategorilere göre NVD_In olarak kodlanmış bir ek şunu ifade etmektedir. Fiilden isim yapan -in veya -un yapım eki, fiile eklenerek onu isme dönüştürmüştür. Ayrıca ¶ sembolü ilgili ekin işlek olduğunu, * sembolü ekin işlek olmadığını ifade etmektedir. Ekin başında parantez ile gösterilen harf, ekin eklendiği gövdeye göre eklenebilir ya da düşebilir olduğunu göstermektedir.

Tablo 2.4: Gramatik kategoriler ve harf kodlaması.

N	Noun	İsim
V	Verb	Fiil
A	Adverb	Zarf
J	Adjective	Sıfat
R	Pronoun	Zamir
P	Postposition	Son çekim edatı
C	Conjunction	Bağlama edatı
X	Exclamation	Ünlem

2.3.2. İsim Çekim Ekleri

Özbekçede kullanılan isim çekim ekleri bu bölümde sunulmuş ve her bir alt bölümde detaylandırılmıştır. Özbek Türkçesinde isim çekim ekleri, çokluk-çoğul ekleri (NNI_PLU), iyelik ekleri (NNI_POSSxx), hâl ekleri (NNI_xxx), aitlik (NNI_REL), eşitlik (NNI_EQU) ve soru çekim ekleri (NNI_QUE) ekleri olup, ek tablosunda detaylı olarak sunulmuştur. Tablo 2.5'te ekler detaylı olarak gösterilmiştir, sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, açıklaması, örnekleri ve örneklerin anlamları sunulmuştur. Morfolojik gösterimde "0" yalın hali, "+0" sembolü ekin eklendiği kelimeden ayrı yazıldığını göstermektedir.

Tablo 2.5: İsim çekim eklerinin morfolojik gösterimi, kodlaması ve örneklendirmesi.

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Açıklama	Örnek	Anlamı
-lar	¶NNI_PLU	NNI_PLU_lär	IC_COGUL_lär	Çoğul Eki	kitoblar, bolalar, toshlar, tog'lar	kitaplar, çocuklar, taşlar, dağlar
-(I)m	¶NNI_POSS1s	NNI_POSS1s_(H)m	IC_IYELIK_BEN_Hm	1.Tekil Şahıs Sahiplik Eki	boshim, bolam	başım, çocuğum
-(I)mIz	¶NNI_POSS1p	NNI_POSS1p_(H)miz	IC_IYELIK_BIZ_Hmiz	1.Çoğul Şahıs Sahiplik Eki	boshimiz, bolamiz	başımız, çocuğumuz
-(I)ng	¶NNI_POSS2s	NNI_POSS2s_(H)ng	IC_IYELIK_SEN_Hng	2.Tekil Şahıs Sahiplik Eki	boshing, bolang	başın, çocuğun
-(I)ngiz	¶NNI_POSS2p1	NNI_POSS2p1_(H)ngiz	IC_IYELIK_SIZ_Hngiz	2.Çoğul Şahıs Sahiplik Eki	boshingiz, bolangiz	başınız, çocuğunuz
-(I)nglar	¶NNI_POSS2p2	NNI_POSS2p2_(H)nglär	IC_IYELIK_SIZ_Hnglär	2.Çoğul Şahıs Sahiplik Eki	boshinglar, bolanglar	başlarınız, çocuklarınız
-laring	¶NNI_POSS2p3	NNI_POSS2p3_läring	IC_IYELIK_SIZ_läring	2.Çoğul Şahıs Sahiplik Eki	boshlaring, bolalaring	başlarınız, çocuklarınız
-(s)I	¶NNI_POSS3s	NNI_POSS3s_(s)H	IC_IYELIK_O_(s)H	3.Tekil Şahıs Sahiplik Eki	boshi, bolasi	başı, çocuğu
-lari	NNI_POSS3p	NNI_POSS3p_läri	IC_IYELIK_ONLAR_läri	3.Çoğul Şahıs Sahiplik Eki	boshlari, bolalari	başları, çocukları
-ni	¶NNI_OBJ	NNI_OBJ_ni	IC_HAL_BELIRTME_ni	Belirtme Hâli	boshni, bolani	başı, çocuğu

Tablo 2.5 (devam):

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Açıklama	Örnek	Anlamı
-n	¶NNI_OBJ3	OBJ3_n	OBJ3_n	Belirtme Hâli (3.Tekil Şahıs Sahiplik Ekinden Sonra)	boshin, bolasin	başını, çocuğunu
0	NNI_NOM	NNI_NOM_0	IC_HAL_YALIN_	Yalın Hâli	bosh, bola	baş, çocuk
-ning	¶NNI_GEN	NNI_GEN_ning	IC_HAL_ILGI_ning	İlgi Hâli	boshning, bolaning	başın, çocuğun
-GA	¶NNI_DAT	NNI_DAT_GA	IC_HAL_YONELME_Ga	Yönelme Hâli	boshga, bolaga, boqqa, terakka, buzoqqa	başa, çocuğa, bahçeye, kavak ağacına, buzağıya
-nga	¶NNI_DAT3	NNI_DAT3_ngä	IC_HAL_YONELME_ngä	Yönelme Hâli (u,bu,şu gibi 3.Tekil Şahıs Sahiplikten Ekinden Sonra)	unga, bunga, shunga	ona, buna, şuna
-Da	¶NNI_LOC	NNI_LOC_dä	IC_HAL_BULUNMA_dä	Bulunma Hâli	boshda, bolada	başta, çocukta
-nda	¶NNI_LOC3	NNI_LOC3_nda	IC_HAL_BULUNMA_nda	Bulunma Hâli (u,bu,şu gibi 3.Tekil Şahıs Sahiplikten Sonra)	unda, bunda, shunda	onda, bunda, şunda

Tablo 2.5 (devam):

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Açıklama	Örnek	Anlamı
-dan	¶NNI_ABL	NNI_ABL_dän	IC_HAL_AYRILMA_dän	Ayrılma Hâli	boshdan, boladan	baştan, çocuktan
-ndan	¶NNI_ABL3	NNI_ABL3_ndän	IC_HAL_AYRILMA_ndän	Ayrılma Hâli (u,bu,şu gibi 3.Tekil Şahıs Sahiplikten Sonra)	undan, bundan, shundan	ondan, bundan, şundan
-+0bilan	¶NNI_INC	NNI_INC_0bilän	IC_HAL_VASITA_bilän	Vasıta Hâli	til bilan, qalam bilan	dille, kalemle
-gi	¶NNI_REL1	NNI_REL1_gi	IC_AITLIK_gi	Aitlik Eki (Ünlülerden, Tonlu Ünsüzlerden ve Bulunma Hâli Ekinden Sonra)	bahorgi, burungi. kechagi, mendagi	baharki, önceki, dünkü, bendeki
-Qi*	NNI_REL2	NNI_REL2_Qi	IC_AITLIK_Qi	Aitlik Eki (Tonsuz Ünsüzlerden Sonra)	ichki, tashqi, dastlabki	içerdeki, dışardaki, ilk önceki
-niki	¶NNI_REL3	NNI_REL3_niki	IC_AITLIK_niki	Aitlik Eki (İlgi Hâli Ekinden Kısalmış -ni Ekinden Sonra)	meniki, uniki, kimniki, kitobniki	benimki, onunki, kiminki, kitabınki
-mI	¶NNI_QUE	QUE_mi	SORU_mi	Soru Eki	Omonmisan, o'qiganmisan	Yaşıyor musun? Okuyor musun?

2.3.2.1. Çokluk Eki

Özbekçede çoğul – çokluk eki, kelime sonuna ve kelime yapısına bakmaksızın tek şekilli olup, -lar ile gösterilir. Örneğin; kitob-lar (kitaplar), maktab-lar (mektepler), bola-lar (çocuklar) ve qiz-lar (kızlar) örneklerinde olduğu gibi tek bir çoğul eki farklı kelimelere gelerek çoğullaştırmıştır.

2.3.2.2. İyelik Ekleri

Özbekçede iyelik ekleri eklendiği kelimenin son harfine göre farklı varyantlarda gösterilebilmektedir. Kelime ile ek arasına sesli harfin eklendiği, türemenin ve tonsuzlaşmanın olduğu örnekler yer almaktadır. Tablo 2.6'da şahıslarına göre ünsüzle biten kelimelerden sonra gelen iyelik ekleri gösterilmiş olup kitob-im (kitabım), maktab-im (mektebim) örneklerinde olduğu gibi iki ünsüz arasında bir ünlü harf olduğu gözlemlenir. Tablo 2.7'de şahıslarına göre ünlüyle biten kelimelerden sonra gelen iyelik ekleri gösterilmiştir. Ünlü bir harften sonra direkt ünsüz bir harf ek olarak eklenebilmektedir. Örneğin, bola-m (çocuğum) ve ona-m (annem) örneklerinde gösterildiği gibi ünlü bir harfle biten kelimeye iyelik eki ünlü almadan direkt eklenir.

Tablo 2.6: Ünsüzle biten kelimelere eklenen iyelik ekleri ve örnekleri.

Tekil ve Çoğul Şahıslar	**İyelik Eki**	**Örnek**
1. Tekil Şahıs	-im	kitobim (kitabım), maktabim (mektebim)
2. Tekil Şahıs	-ing	kitobing (kitabın), maktabing (mektebin)
3. Tekil Şahıs	-i	kitobi (kitabı), maktabing (mektebi)
1. Çoğul Şahıs	-imiz	kitobimiz (kitabımız), maktabimiz (mektebimiz)
2. Çoğul Şahıs	-ingiz, -inglar, -laring	kitobingiz (kitabınız), maktabingiz (mektebiniz)
3. Çoğul Şahıs	-i, -lari	kitoblari (kitapları), maktablari (mektepleri)

Tablo 2.7: Ünlüyle biten kelimelere eklenen iyelik ekleri ve örnekleri.

Tekil ve Çoğul Şahıslar	**İyelik Eki**	**Örnek**
1. Tekil Şahıs	-m	bolam (çocuğum), onam (annem)
2. Tekil Şahıs	-ng	bolang (çocuğun), onang (annen)
3. Tekil Şahıs	-si	bolasi (çocuğu), onasi (anası)
1. Çoğul Şahıs	-miz	bolamiz (çocuğumuz), onamiz (anamiz)
2. Çoğul Şahıs	-ngiz, -nglar, -laring	bolangiz (çocuğunuz), onangiz (anneniz)
3. Çoğul Şahıs	-si, -lari	bolalari (çocukları), onalari (anneleri)

Bazı kelimelere eklenen iyelik ekleri türemeye, bazı kelimelerde tonlulaşmaya neden olur. Örneğin, parvo-yi-miz (önemsememiz) ve obro'-y-ing (itibarın) örneklerinde y ve yi türemeleri gösterilmektedir. Son sesleri k veya q ünsüzleri olan çok heceli kelimelere ve bazı tek heceli kelimelere (bek (bey), yo'q (yok) vb.) iyelik eki geldiğinde kelimelerin sondaki tonsuz ünsüzleri tonlulaşarak k-g ve q-g' ses değişimleri meydana gelir ve imlâda gösterilir. Örneğin, tilak-ing kelimesinde k-g tonlulaşması gözlemlenerek tilaging (dileğin) ve qoshiq-i kelimesinde q-g' tonlulaşması gözlemlenerek qoshig'i (kaşığı) olarak yazılmaktadır.

2.3.2.3. İsim Hâl Ekleri

Özbekçede ismin altı farklı hâl eki (yalın, ilgi, yönelme, yükleme, bulunma ve çıkma hâl ekleri) bulunmaktadır. Bu hâl eklerinde ses düşmelerinin görülmesinin yanında, yönelme hâl ekinin ses değişiminden dolayı farklı şekilleri yer almaktır.

Son hecesi k ve q sesleriyle biten kelimelere yönelme hâl eki -ga geldiğinde ek başındaki g sesi kelime sonundaki ses tarafından değiştirilir ve imlâda gösterilir. Bu durum terak-ga ve buzoq-ga örneklerinde gösterildiği üzere ses değişimi sonrası bu kelimeler terakka (kavak ağacına) ve buzaqqa (buzağıya) olarak yazılır. İsmin hâlleri, hâl çekim ekleri ve örnekleri Tablo 2.8'de gösterilmiştir.

Tablo 2.8: İsim hâl ekleri.

Hâller	**Hâl Çekim Ekleri**	**Örnekler**
Yalın Hâl	-	kitob (kitap)
İlgi Hâli	-ing, -ning	kitobning (kitabın)
Yönelme Hâli	-ga, -nga, -ka, -qa	kitobga (kitaba)
Yükleme Hâli	-ni, -i	kitobni (kitabı)
Bulunma Hâli	-da, -nda	kitobda (kitapta)
Çıkma Hâli	-dan, -ndan	kitobdan (kitaptan)

2.3.2.4. Soru Eki

Özbek Türkçesinde soru eki sadece bir tane olup, -mi olarak yazılır. Soru eki, şahıs eklerinden sonra gelir. Bu kural her zaman geçerli olmayıp, istisnaları mevcuttur ve imlâda gösterilir. Şahıs eklerinden önce gelerek, isim tabanından ya da çekimli fiilden hemen sonra gelebilir ve bitişik yazılır. Bu durum omon-mi-san (yaşıyor musun?) örneğinde gösterilmiştir.

2.3.2.5. Olumsuzluk Eki

Özbekçede olumsuzluk anlamı, fiillerde -ma olumsuzluk ekiyle, isimlerde emas (değil) ve yoq (yok) gibi olumsuzluk edatlarıyla ya da -siz olumsuzluk eki ile yapılır. Ancak, bazı fiil çekimlerinde de olumsuzluk edatları kullanılabilir. Bu durumlar borgan yoq (gitmedi), olganim yok (almadım) ve kelgan emas (gelmedi) gibi örneklerde gösterilmiştir.

2.3.3. İsim Yapım Ekleri

Özbek Türkçesinde kullanılan isim yapım ekleri, bu bölümde anlatılarak her biri alt bölümde detayları ve ek tabloları ile paylaşılmıştır. Özbekçede isim yapım ekleri, isimden isim yapan (NND_xxxxxx), sıfattan isim yapan (NJD_xxxx), fiilden isim yapan (NVD_xxxxxx), fiilden sıfat yapan (JVD_xxxx), sıfattan sıfat yapan (JJD_xxxxx), isimden sıfat yapan (JND_xxxxx) ve zarf yapan (AND_xxxx, AJD_xxxxxx ve AAD_xxxxx) ekler olup, alt bölümlerde detaylandırılmıştır.

2.3.3.1. İsimden İsim Yapan Yapım Ekleri

Özbekçe'de isimlerden isim yapan yapım ekleri eklendikleri isim kök ve gövdelerinden yeni isim türeten yapım ekleridir. Aşağıdaki tabloda bu ekler sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, örnekleri ve örneklerin anlamlarıyla detaylı sunulmuştur.

Tablo 2.9: İsimden isim yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	**Kodu**	**Nuve Studio Kodu (İngilizce)**	**Nuve Studio Kodu (Türkçe)**	**Örnek**	**Anlamı**
-a*	NND_A	NND_ä	IY_ISIM_ä	erka, beka, yuza	şımarık, asil hanım, yüzey
-cha	¶NND_CHA	NND_çä	IY_ISIM_çä	Turkcha, Latincha, qopcha, qizcha, kitobcha	Türkçe, Latince, küçük torba, kızcağız, kitapçık
-chi	¶NND_CHI	NND_çi	IY_ISIM_çi	yolg'onchi, qiziqchi, sovchi, tankidchi, o'raqchi	yalancı, şakacı, dünür, tenkitçi, orakçı
-chIQ	¶NND_CHIQ	NND_çIQ	IY_ISIM_çIQ	irgamchik, qapchik, to'ychiq, o'yinchiq, kepchick	iğrenç, cüzdan, küçük düğün, oyuncak, yelpaze
-chAQ	¶NND_CHAQ	NND_çAQ	IY_ISIM_çAQ	kelinchak, qo'zinchoq, cho'hoq	gelinceğiz, kuzucuk, çalı çırpı
-chAN	¶NND_CHAN	NND_çAn	IY_ISIM_çAn	ishchan, uyatchang	havalı, utangaç
-AQ*	NND_AQ	NND_AQ	IY_ISIM_AQ	boshoq, ichak	başak, bağırsak
-dosh	¶NND_DOSH	NND_dåş	IY_ISIM_dåş	tengdosh, yo'ldosh, asrdosh, maslakdosh	yaşıt, yoldaş, zamandaş, meslektaş
-GInä	¶NND_GINA	NND_GHnä	IY_ISIM_GHnä	bilakkina	Bilekcik
-loq	¶NND_LOQ	NND_låq	IY_ISIM_låq	yayloq, qishloq, toshloq, qumloq, o'tloq, tuzloq	yayla, kışla, taşlı yer, kumlu yer, otlak, tuzlu yer
-DIq*	NND_DIQ	NND_DIq	IY_ISIM_DIq	burundiq, orindiq, qoltiq	burunluk, sandalye, koltuk

Tablo 2.9 (devam):

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-Ga*	NND_GA	NND_Ga	IY_ISIM_Ga	yo'lka, o'zga, boshqa,	tratuvar, özge, başka
-lIQ	¶NND_lIQ	NND_lIQ	IY_ISIM_lIQ	nomozlik, tuzluq, kuyovlik	seccade, tuzluk, damatlık
-mån*	NND_MON	NND_mån	IY_ISIM_mån	bilormon, ustomon	bilici, açgöz

2.3.3.2. Sıfattan İsim Yapan Yapım Ekleri

Sıfattan isim yapan yapım ekleri, sıfat kök ve gövdelerine gelerek isim türeten yapım ekleri olup, aşağıdaki tabloda bu ekler sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, örnekleri ve örneklerin anlamlarıyla detaylı sunulmuştur.

Tablo 2.10: Sıfattan isim yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-cha	¶NJD_CHA	NJD_çä	SY_ISIM_çä	qizilcha, ko'kcha, olacha	kırmızı pancar, kavun çeşidi, çizgili kumaş
-chIQ	¶NJD_CHIQ	NJD_çHQ	SY_ISIM_çHQ	qorachiq	gözbebeği
-lIQ	¶NJD_LIQ	NJD_lIQ	SY_ISIM_lIQ	ko'plik, ozlik, yaxshilik	çokluk, azlık, iyilik

2.3.3.3. Fiilden İsim Yapan Yapım Ekleri

Fiilden isim yapan yapım ekleri, fiil kök ve gövdelerine eklenerek isim yapan yapım ekleri olup, aşağıdaki tabloda bu ekler sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, örnekleri ve örneklerin anlamlarıyla detaylı sunulmuştur.

Tablo 2.11: Fiilden isim yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-(A)Q	¶NVD_AQ	NVD_(A)Q	FY_ISIM_AQ	bo'lak, qo'naq, qochoq	bölük, konuk, kaçak
-(V_{bo})v	¶NVD_OV	NVD_Hv	FY_ISIM_Hv	sinov, to'lov, yazuv, surov	sınav, ücret, yazı, soru
-(In)ch	¶NVD_INCH	NVD_Hnç	FY_ISIM_Hnç	o'kinch, qizg'onch, tayanch, sevinch, qorqunch, tiqilinch	pişmanlık, kıskanç, sağlam, sevinç, korkunç, tıkanık
-chIq*	NVD_CHIQ	NVD_çHq	FY_ISIM_çHq	suvanchiq, sirg'anchiq, qizg'onchiq	baston, kaygan, kıskanç
-dAQ	¶NVD_DOQ	NVD_dÅQ	FY_ISIM_dAQ	kekirdak, qovurdoq	gırtlak, kavurma
-dIQ	¶NVD_DIQ	NVD_dHQ	FY_ISIM_dHQ	siydik, tashlandiq, qoldiq, topildiq	idrar, lüzumsuz, atık, buluntu
-GI	¶NVD_GI	NVD_GH	FY_ISIM_GH	yoqilg'i, kulgi, sanchqi, tuyg'u, turtki	yakıt, gülüş, çatal, duygu, dürtü
-GIch	¶NVD_GICH	NVD_Giç	FY_ISIM_GHç	purkagich, yelpig'ich, ko'rsatkich, yulg'ich	püskürtücü, yelpaze, gösterge, menfaatçi
-GIn	¶NVD_GIn	NVD_GHn	FY_ISIM_GHn	sotgin, turg'un, tushkun	hayin, durgun, düşkün
-(I)Q	¶NVD_IQ	NVD_IQ	FY_ISIM_IQ	kekirik, uzuk, ag'riq, yumuq	geğirti, yüksük, ağrı, yumuk
-(GI)nDI	¶NVD_GINDI	NVD_GINDI	FY_ISIM_GINDI	kelgindi, yig'indi, yuvindi	yabancı, yığın, bulaşık suyu
-(I)m	¶NVD_IM	NVD_Im	FY_ISIM_Im	siqim, tutam, yutum	sıkım, bir tutam, yudum
-ma	¶NVD_MA	NVD_mä	FY_ISIM_mä	bo'lma, bosma, yozma	oda, baskı, yazılı

Tablo 2.11 (devam):

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-moq	¶NVD_MOQ	NVD_ måq	FY_ISIM_måq	qarmoq, ilmoq, chakmoq	olta, sürgü, tebessüm etmek
-(I)n*	NVD_IN	NVD_Hn	FY_ISIM_Hn	bog'in, yog'in, yig'in	hece, yağış, toplantı
-(I)sh	¶NVD_ISH	NVD_(H)ş	FY_ISIM_Hş	urush, urish, garg'ish, burilish	savaş, vuruş,kargış, kıvrım
-mIr	¶NVD_MIR	NVD_mIr	FY_ISIM_mIr	o'smir, yomg'ur, yalamur	ergen çocuk, yağmur, açgözlü
-mIsh	¶NVD_MISH	NVD_mHş	FY_ISIM_mHş	o'tmish, kechmish, turmush, yemish	geçmiş, geçmiş, hayat, yemek
-(on)g'Ich	¶NVD_ONGICH	NVD_ong'Ich	FY_ISIM_ong'Ich	tepong'ich, suzong'ich, qopog'ich	devamlı tepen, devamlı toslayan, ısırgan

2.3.3.4. Fiilden Sıfat Yapan Yapım Ekleri

Özbekçede fiilden sıfat yapan yapım ekleri, eklendiği fiil kök ve gövdelerinden sıfat türeten yapım ekleri olup, aşağıdaki tabloda bu ekler sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, örnekleri ve örneklerin anlamlarıyla detaylı sunulmuştur.

Tablo 2.12: Fiilden sıfat yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-ag'on*	JVD_AG'ON	JVD_äġån	FY_SIFAT_äġån	bilag'on	çok bilen
-asi	JVD_ASI	JVD_äsi	FY_SIFAT_äsi	o'lasi, kelasi	ölesiye, gelecek

Tablo 2.12 (devam):

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-(E)nch	¶JVD_ENCH	JVD_(H)Nç	FY_SIFAT_Hnç	qizg'anch, qorginch	kıskanç, korkunç
-chAQ	¶JVD_CHAQ	JVD_çÅQ	FY_SIFAT_çAQ	erinchak, tortinchoq	uyuşuk, çekingen
-QAQ	¶JVD_QAQ	JVD_QÅQ	FY_SIFAT_QAQ	urishqoq	Kavgacı
-chIQ	¶JVD_CHIQ	JVD_çHQ	FY_SIFAT_çHQ	quvonchik, qizg'anchik,qavarchik	neşeli, kıskanç, kabarcık

2.3.3.5. Sıfattan Sıfat Yapan Yapım Ekleri

Sıfattan sıfat yapan yapım ekleri, sıfat kök ve gövdelere gelerek sıfat türeten ekler olup, az sayıda olmasına rağmen sıfattan türetilen isimler ile karıştırılmaması için ayrıştırılarak ayrıca aşağıdaki tabloda sunulmuştur.

Tablo 2.13: Sıfattan sıfat yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-cha	¶JJD_CHA	JJD_çä	SY_SIFAT_çä	qoracha, ko'kcha, yangicha, olacha	karamsı, mavimsi, yenice, alaca
-GIna	¶JJD_GINA	JJD_GHnä	SY_SIFAT_GHnä	sekingina, quruqqina	sakince, kupkuru
-Imtir	¶JJD_IMTIR	JJD_Hmtir	SY_SIFAT_Hmtir	oqimtir, qoramtir, sarg'imtir, qizg'imtir	akan, koyu, sarımsı, kırmızımsı
-roq	¶JJD_ROQ	JJD_råq	SY_SIFAT_råq	yaxshiroq, eskiroq, pastroq	daha iyi, daha yaşlı, daha düşük
-ish	¶JJD_ISH	JJD_iş	SY_SIFAT_iş	ko'kish	Kokuşmuş,
-g'ish	¶JJD_G'ISH	JJD_g'iş	SY_SIFAT_g'iş	sarg'ish	sarımsı, kırmızımsı
-(i)nchi	¶JJD_INCHI	JJD_Inçi	SY_SIFAT_Inçi	oltinchi	altıncı

2.3.3.6. İsimden Sıfat Yapan Yapım Ekleri

Özbekçede isimden sıfat yapan yapım ekleri, isim kök ve gövdelere gelerek sıfat türeten yapım ekleri olup, aşağıdaki tabloda bu ekler sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, örnekleri ve örneklerin anlamlarıyla detaylı sunulmuştur.

Tablo 2.14: İsimden sıfat yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	**Kodu**	**Nuve Studio Kodu (İngilizce)**	**Nuve Studio Kodu (Türkçe)**	**Örnek**	**Anlamı**
-chIl	¶JND_CHIL	JND_çil	IY_SIFAT_çil	dardchil, izchil, kamchil, kirchil	dertli, tutarlı, az bulunmaz, çabuk kirlenen
-li	¶JND_LI	JND_li	IY_SIFAT_li	kitapli, dimog'li, hurmatli, chiroyli	kitaplı, kibirli, hürmetli, güzel
-lIQ	¶JND_LIQ	JND_lHQ	IY_SIFAT_lHQ	ko'plik, ozlik, yaxshilik	çokluk, azlık, iyilik
-simon	¶JND_SIMON	JND_simån	IY_SIFAT_simån	oysimon, ignasimon	ay gibi, iğne biçimli dendrit
-siz	¶JND_SIZ	JND_siz	IY_SIFAT_siz	vafosiz, keraksiz, uyqusiz	vefasız, gereksiz, uykusuz
-ta	¶JND_TA	JND_tä	IY_SIFAT_tä	beshta, oltita, bitta	beş tane, altı tane, bir tane
-(v)iy	¶JND_VIY	JND_(v)iy	IY_SIFAT_(v)iy	adabiy, partiyaviy, irqiy, irsiy, insonly	edebi, partizan, ırksal, kalıtsal, insan

2.3.3.7. Zarf Yapan Yapım Ekleri

İsim, sıfat ve zarf gibi isim soylu kök ve gövdelere eklenerek zarf türeten yapım ekleridir. Aşağıdaki tabloda sırasıyla ekin biçimbilimsel gösterimi, kodlaması, Nuve Studio İngilizce ve Türkçe kodlaması, örnekleri ve örneklerin anlamlarıyla detaylı sunulmuştur.

Tablo 2.15: Zarf yapan yapım eklerinin morfolojik gösterimi ve örneklendirmesi.

Biçimbilimsel Gösterim	Kodu	Nuve Studio Kodu (İngilizce)	Nuve Studio Kodu (Türkçe)	Örnek	Anlamı
-(n)cha	¶AJD_NCHA	JND_(v)iy	IY_SIFAT_(v)iy	buncha, mardcha, o'zlaricha, meningcha, o'zicha, bizningcha	bunca, bu kadar, kendilerince, bence, kendince, bizce

2.4. İKİ DÜZEYLİ BİÇİMBİLİMSEL (MORFOLOJİK) ÇÖZÜMLEME

Bir dilin morfolojik çözümlemesi kelimelerin aldığı eklerin ayrışarak kelime gövdesine ve en yalın hali olan kök haline ayrıştırılmasının ve kelime türlerinin (isim, sıfat, fiil, zarf vb.) belirtilmesi işlemidir.

İki düzeyli morfoloji metodu (Koskenniemi, 1983) için bir sözcüğün iki gösterimi vardır. Bunlar sözcüksel (lexical) ve yüzeysel (surface) gösterim biçimleridir. Sözcüksel (lexical) biçim, bir sözcüğün işlevsel bileşenlerinin yapısını belirtir ve gösterir. Sözcüksel biçim sözcüğün yapısı ile alakalı olduğu için yapısal biçim olarak ayrıca ifade edilmektedir. Bir sözcüğün hangi morfolojik yapılardan(morfemlerden) oluştuğunun, sözcüğün içerdiği eklerin (çekim eki-yapım eki gibi) ve özellikleri ile türlerinin yanı sıra eklenme sıraları ile çözümlendiği tüm sözcük yapısının gösterildiği biçimdir. Yüzeysel (surface) biçimi, sözcük diziminin, sözcük yapısının standart yazım kuralları ile gerçekleştirilmesini göstermektedir. Sözcüğün yüzeysel gösterimi metin içerisinde görünen imlasal biçimidir. Sözcüğün çekim – yapım eklerini aldıktan sonraki halini, yani metinde eklerin eklenmiş haldeki sözcüğün gösterim biçimini ifade etmektedir.

İki düzeyli morfolojik çözümlemede, kelimelerin aldığı ekler ve kelime kökü ayrıştırılırken yazım kuralları (ortografik kurallar) ve kelimelerin ekler aldığı zamanlarda ortaya çıkan yazım değişiklikleri göz önünde bulundurulur. Harf ve hece boyutundaki değişmeler ve düşmeler iki düzeyli yazım kuralları olarak tanımlanır. Böylece alınan eklerin kelimede veya ekte meydana getirdiği değişikliklerde çözümleme esnasında belirlenmiş olur. Biçim dizgeleri ise kelimeye hangi eklerin (çekim-yapım) gelebileceğini ve bu eklerin hangi sırada yer aldığını gösterir. Biçim dizgeleri (morfotaktik) ise sonlu durum makineleri (FSM) kullanılarak tanımlanır.

İki düzeyli model ile ifade edilen, sözcüksel ve yüzeysel biçimler arasındaki değişimler yazım kurallarına göre aşağıda ifade edilen dört ana kuralla ifade edilmektedir.

1) a:b => LC__RC

Bu kurala göre, sözcüksel (lexical) a'nın, yüzeysel (surface) b'ye tekabül etmesi sadece sözcüksel a'nın kendinden önce (LC-Left Context) ve sonra (RC-Right Context) belirtilen bağlamları sağladığında gerçekleşeceğini belirtir. Bu dönüşüm sadece bu bağlamda gerçekleşir, daima gerçekleşmez.

2) a:b <= LC__RC

Bu kurala göre, sözcüksel (lexical) a'nın, yüzeysel (surface) b'ye dönüşümü kesin olarak sözcüksel a'nın kendinden önce (LC-Left Context) ve sonra (RC-Right Context) belirtilen bağlamları sağladığında gerçekleşeceğini ve farklı durumlarda da gerçekleşebileceğini belirtir. Bu dönüşüm kesinlik belirtmektedir, daima gerçekleşir ve bu durum haricindeki durumlarda da gerçekleşebilir.

3) a:b <=> LC__RC

Bu kurala göre, sözcüksel (lexical) a'nın, yüzeysel (surface) b'ye dönüşümü kesin olarak sözcüksel a'nın kendinden önce (LC-Left Context) ve sonra (RC-Right Context) belirtilen bağlamları sağlandığında gerçekleşeceğini belirtir. Bu dönüşm bu bağlamda kesinlik belirtmektedir. Daima gerçekleşir ancak bu durum haricindeki durumlarda da asla gerçekleşmez.

4) a:b /<= LC__RC

Bu kurala göre, sözlüksel (lexical) a'nın, yüzeysel (surface) b'ye dönüşümü kurala bağlı olarak gerçekleşmez. Bu kural yapısal a'nın kendinden önce (LC-Left Context) ve sonra (RC-Right Context) belirtilen bağlamları kuralı sağladığı durumdur. Bu bağlamda dönüşüm kesinlikle gerçekleşmez.

İki düzeyli kuralları tanımlarken bazı semboller kullanılmaktadır. Kullanılan bu semboller ve anlamları şunlardır:

- * önceki harfin sıfır ya da daha fazla tekrarlanması,

- **?** önceki harfin sıfır ya da bir kez tekrarlandığı,
- **0** yüzeyde görünmediği,
- **@** kelime sonunu,
- **$** sonraki harfin yüzeyde düştüğü,
- + ek sınırlarını gösterirken,
- **[a | b | c]** ya da **[abc]** ise a veya b veya c,

Yukarıdaki semboller ve her bir sembolün yanındaki açıklama sembolünü anlamını ve kullanım amacını açıklamaktadır.

2.5. BİÇİMBİLİMSEL BELİRSİZLİK GİDERME

Bir dilin biçimbilimsel çözümlemesi sonrası, bir kelimenin birden fazla biçimbilimsel çözümü olabilmektedir. Cümlenin içerisindeki bir kelimenin o cümle içerisindeki en doğru biçimbilimsel çözümünün bulunması için uygulanan yöntemler biçimbilimsel belirsizlik gidermenin konusudur.

Bir kelimenin biçimbilimsel çözümlemesini üreten biçimbilimsel çözümleyicilerin ürettiği tüm çözümler gramer olarak doğru olup, birden fazla olabilir. Bu çözümlerin herhangi birinin yanlış olduğu anlamına gelmez. Ancak, bu çözümlerin hangisinin o cümle için doğru olduğu belirsizliğin giderilmesi gerekmektedir. Bu kapsamdaki çalışmalar neticesinde tek bir çözüm en doğru çözüm olarak gösterilmektedir.

Bu çalışmada Dayanık ve diğ. (2018) çalışması referans alınarak Özbekçe için biçimbilimsel belirsizlik giderme MorphNet modeli kullanılarak araştırılmıştır. MorphNet girdi olarak aldığı metnin içerisindeki kelimelerin doğru biçimbilimsel çözümünü üreten, bir sıralı (seq2seq) tekrarlayan yapay sinir ağı (RNN) modelidir. MorphNet her ne kadar girdi olarak aldığı metinin içerisindeki bir kelimenin biçimbilimsel çözümleme ve sonrasında belirsizlik giderme işlemini gerçekleştirse de bu çalışma kapsamında sadece biçimbilimsel belirsizlik giderme işlemleri için kullanılmıştır.

2.6. SÖZDİZİMSEL ÇÖZÜMLEME

Bir dilin sözdizimsel çözümlemesi, verilen bir cümlenin dil bilgisi kurallarına uygun bir şekilde sözdiziminin gerçekleştiğinin doğrulandığı ve cümleyi oluşturan her bir morfolojik öğenin bu

kurallara uygun dizildiğinin kontrolü, karşılaştırılarak ölçümlenmesi ve ayrıştırılması demektir. Verilen bir cümle daha küçük parçalara ayrıştırılarak kurallara uygunluğu kontrol edilir, her bir cümlenin çözümlenmesi için cümleyi oluşturan kelimelerin ve kelime gruplarının çözümlenmesi gerekmektedir. Kelimelerin birbirleriyle olan ilişkilerinin analiz edilerek kelime türlerinin belirlenmesi gerekmektedir. Bir cümlenin sözdizimsel çözümlemesi için cümleyi oluşturan kelimelerin biçimbilimsel (morfolojik) çözümlemesine ihtiyaç duyulmaktadır. Böylece kelime ve kelimenin ekleri (morfemleri) ile cümle içerisindeki kelimenin türü ve cümlenin hangi ögesi olduğu çözümlenmiş olacaktır.

Herhangi bir dilin sözdizimsel çözümlemesinin yapılması için öncelikli olarak o dilin gramerinin ve sözdiziminin detaylı bir şekilde incelenmesi gerekmektedir. Cümle ögelerinin sırası, sıralanışı ve cümle içindeki yeri ile cümle yapıları araştırılmalıdır. Dil kurallarına uygun geçerlilik-doğrulama kuralları tanımlanmalıdır. Böylece sözdizimi kontrolü sağlanacaktır. Ayrıca cümleyi oluşturan kelime ve kelime gruplarının çözümlenmesi gerekmektedir. Kelime ve kelime gruplarının birbiri ile olan ilişkilerinin de çözümlenmesi sözdizimsel çözümlemenin bir parçasıdır.

Sözdizimsel çözümleme için “Yukarıdan – Aşağıya (Top – Down)” sözdizimsel çözümleyici yaklaşımına sahip, morfolojik çözümlemeye, bağlamdan bağımsız gramer kurallarına ve geçerlilik-doğrulama kurallarına dayanan bir çalışma referans (Zafer, 2011) alınarak bu çalışmadaki yöntem kullanılmıştır. Cümle türleri ve kelime grupları çalışılarak dil gramer kuralları tanımlanmıştır. Sözdizimsel çözümleme çıktıları olası tüm çözümleri sunarak cümleyi bileşenlerine ayırmış ve bir ağaç yapısı ile ifade etmiştir.

3. MALZEME VE YÖNTEM

Bu bölümde Özbekçenin biçimbilimsel çözümlemesi, biçimbilimsel belirsizlik giderme ve sözdizimsel çözümlemesi için önerilen ve uygulanan yöntemler anlatılarak, alt bölümlerde detaylandırılacaktır.

Özbekçenin biçimbilimsel çözümlemesi alt bölümünde, yukarıda genel tanımlamaları yapılan biçimbilimsel çözümleme üzerinde durularak, iki düzeyli morfoloji ve iki düzeyli yazım kuralları anlatılmış, biçim dizgesi kuralları tanımlanarak, isim ve fiil gövdeleri biçim dizgesi kuralları detaylandırılmıştır. Biçimbilimsel belirsizlik giderme bölümünde, biçimbilimsel çözümleme bölümünde üretilen çıktılar kullanılarak, belirsizlik durumları gösterilecektir. Belirsizlik durumlarının çözümlemesi ve tek bir çıktının üretilmesi için kullanılan yöntem ve detayları bu alt bölümde yer almaktadır.

Özbekçenin sözdizimsel çözümlemesi üzerine yapılan araştırmalar, çalışmalar, geliştirmeler ve uygulanan yöntem bu bölümün son alt bölümünde detaylandırılmıştır. Özbekçe dil grameri detaylı bir şekilde incelendikten sonra sözdizimsel çözümleme için kullanılan yöntem belirlenmiştir. Buna göre Özbekçenin sözdizimsel çözümlemesi için "Yukarıdan – Aşağıdaya (Top-Down)" sözdizimsel çözümleyici yaklaşımına sahip, biçimbilimsel çözümlemeye, bağlamdan bağımsız gramer kurallarına ve geçerlilik-doğrulama kurallarına dayan bir yöntem uygulanmıştır.

3.1. ÖZBEKÇENİN BİÇİMBİLİMSEL ÇÖZÜMLEMESİ

Biçimbilimsel (morfolojik) çözümleme bir kelimeyi oluşturan gövde ve eklerin ayrıştırılarak, gövdenin en yalın hali olan kök haline ayrıştırılması ve kelimeyi oluşturan bileşenlerin türleri ile belirtilmesi işlemidir. Biçimbilimsel çözümlemenin yapılması için ilgili dilin yazım (ortografik) kurallarının tanımlanması gerekmektedir. Yazım (ortografik) kuralları bir dilde eklerin gövdeye eklenme biçimi ve sırasını belirleyen kurallar olup, kelimeler bu kurallar doğrultusunda oluşturularak, bir kelime kök haline çözümlenirken bu kurallar doğrultusunda çözümlenmektedir.

Özbekçenin biçimbilimsel (morfolojik) çözümlemesi için bu bölümde iki düzeyli morfoloji (biçimbilimi) ve iki düzeyli ortografi (yazım) kuralları detaylı olarak anlatılmıştır. İki düzeyli morfoloji (biçimbilimi) metodu (Koskenniemi, 1983) genel bir metot olup bu çalışmada

referans alınmıştır. Bu modele göre dil kuralları ve ilgili dile ait sözlük birleştirilerek bir çözümleyici ile herhangi bir dilin çözümlemesi söz konusudur. Bir dilin morfolojisi (biçimbilimi) bu yaklaşımla modellenir ve çözümlenir. İki düzeyli biçimbilimi metoduna göre sözcüksel (lexical) ve yüzeysel (surface) gösterimi ile bir kelimenin iki farklı gösterimi mevcuttur. Sözcüksel (lexical) gösterim, bir sözcüğün hangi morfolojik yapılardan (morfemlerden) oluştuğunun, sözcüğün içerdiği eklerin (çekim – yapım eki gibi) özellikleri ve türleri ile eklenme sıralarını da göstererek tüm sözcük yapısını göstermektedir. Yüzeysel (surface) gösterim, sözcüğü oluşturan ekler ile sıralanışının ve sözcük yapısının yazım kurallarına uygun bir şekilde, cümle içerisinde yazıldığı gibi gösterildiği biçimdir.

İki düzeyli morfolojik çözümlemede, kelimelerin aldığı ekler ve kelime kökü ayrıştırılır. Bu ayrıştırılma yapılırken yazım kuralları dikkate alınır. Kelimelerin ekler aldığı zamanlarda ortaya çıkan yazım değişiklikleri, harf ve hece boyutundaki değişmeler ve düşmeler iki düzeyli yazım kuralları ile ifade edilir. Bu sayede biçimbilimsel çözümleme yapılırken kökün aldığı eklerin kelimede veya ekte meydana getirdiği değişimler belirlenmiş olur. İki düzeyli biçimbilimde yazım kuralları iki düzeyli kurallar ile tanımlanmaktadır. Bu bölümün alt bölümlerinde iki düzeyli Özbek morfolojisi ve Özbekçenin iki düzeyli yazım kuralları detaylandırılmaktadır.

Morfotaktik (biçim dizgesi) kelimeyi oluşturan morfemlerin sıralanışı ve birbirleri ile olan ilişkilerini düzenleyen kurallar dizisidir. Morfotatik kuralları sonlu durum makineleri (finite state machine) ile gösterilmektedir. Bu bölümün alt bölümlerinde Özbekçenin biçimbilimsel çözümlemesi için morfotaktik kuralları detaylandırılmış olup isim ve fiil gövdeleri için kurallar oluşturularak modellenmiş ve sonlu durum makineleri ile gösterilmiştir.

Bu bölümün ilgili alt bölümlerinde Özbekçenin iki düzeyli morfolojisinin detaylandırılması, iki düzeyli yazım kuralları ve morfotaktik kuralları anlatılarak, Özbekçenin biçimbilimsel çözümlenmesi araştırılmıştır.

3.1.1. Özbekçenin İki Düzeyli Yazım Kuralları

Özbekçenin biçimbilimsel çözümlemesinin yapılabilmesi için Özbekçe grameri ve yazım kuralları detaylı bir şekilde araştırılmış olup, Özbekçe gramerine uygun iki düzeyli yazım kuralları tanımlanmıştır. Özbekçe, Türk dilleriyle ünlü-ünsüz düşmesi, ünlü-ünsüz türemesi ve ünlü-ünsüz değişimi gibi konularda benzerlik gösterir. Birçok konuda Türk dilleriyle benzerlik

gösteren Özbekçede düzenli bir ünlü ve ünsüz uyumu bulunmamaktadır (Üşenmez, 2012). Bu tezde Özbekçe dil ve yazım kuralları büyük bir titizlikle incelenerek, iki düzeyli yazım kuralları Özbek morfolojisini tanımlamak için kullanılmıştır. Yazım kurallarının daha sade ve kolay tanımlanması için harf kümelerinden oluşan meta harfler kullanılmıştır. Tanımlamalarda kullanılan meta harfler aşağıdaki gibidir.

C={b,d,f,g,h,j,k,l,m,n,p,q,r,s,t,v,z,y,x,g’,sh,ch,ng},

C_s={f,s,t,k,q,ch,sh,l,n,s,h,y},

C_y= {b, d, g, h, j, m, p, r, v, z, x, g’, ng},

D={d,t},

E={a,i},

G={q,k,g,g’},

I={u,i},

Q={q,k},

V={a,e,i,u,o,o’},

V_{bo}={o,u},

V_{de}={o,i,u},

V_{fu}={a,e,i},

Y={a,y}.

Özbekçe iki düzeyli yazım kuralları, kural açıklaması ve kurala ait örnekler ile sunulmuştur. Her bir kuralda tanımlanan örneklerin ilk satırında örnek kelimenin sözcüksel biçimde gösterimi, ikinci satırında yüzeysel biçimde gösterimi sağlanmıştır. Yazım kurallarına uygun olarak sözcüksel biçiminden yüzeysel biçime dönüşümü gösterilmiş olup, Özbekçe için 26, iki düzeyli kural tanımlanmıştır.

1. Bu kurala göre gövdeye eklenen ekteki ünlü, gövdenin son ünlüsüne göre değişim göstermektedir. Buna göre, gövde son ünlüsü e veya i ise ve bu gövde A={a, o} meta harfini içeren bir ek alırsa, A yüzeyde a olarak gösterilir. Gövde son ünlüsü o veya u ise A yüzeyde o olarak gösterilir. Bu kural için bazı istisnalar bulunmaktadır. Aşağıda, bu+(n)daqa → bundaqa ve shu+(n)daqa → shundaqa örneklerinde bu istisnalar gösterilmektedir.

   A:a => V:[e | i]C*+C*__
   A:o => V:[o | u] C*+C*__

ovun+chAQ	V(teselli ol)+JVD_CHAQ
ovun0choq	
ovunchoq	avuntu
kekir+dAQ	V(geğir)+NVD_DAQ
kekir0dak	
kekirdak	gırtlak
bu+ndAQA	A(bu)+AAD_NDAQA
bu0ndaqa	
bundaqa	böyle
shu+ndAQA	A(shu)+AAD_NDAQA
shu0ndaqa	
shundaqa	şöyle

2. Son ünlüsü o, i veya u olan bir gövdeye E={a, i} meta harfini içeren bir ek gelirse, E yüzeyde i olarak gösterilir. Son ünlüsü o' olan bir gövde, E meta harfini içeren bir ek alırsa E yüzeyde a olarak gösterilir.

   E:i => V:[o | i | u]C*+C*__
   E:a => V:o'C*+C*__

uch+Er	V(uç)+CAUSR_ER
uch0ir	
uchir	birisini / bir şeyi uçurmak
o't+Ger	V(geç)+CAUSR_ER
o't0kar	
o'tkar	bir şeyi / birisini geçirmek

3. Bu kurala göre son ünlüsü u veya e olan bir gövdeye I={u, i} meta harfini içeren bir ek gelirse, I yüzeyde u olarak gösterilir. Gövdenin son ünlüsü i olur ve I meta harfini içeren ek alırsa I yüzeyde i olarak gösterilir.

I:u => V:[e | u]C*+C*__
I:i => V:iC*+C*__

kekir+IQ	V(geğir)+NVD_IQ
kekir0ik	
kekirik	geğirme
uz+IQ	V(kopar)+ NVD_IQ
uz0uq	
uzuq	kopuk
yet+IQ	V(yet)+ JVD_IQ
yet0uk	
yetuk	yetişkin

4. Son ünlüsü a olan bir gövdeye, AQ içeren bir ek geldiğinde, ek yüzeyde aq olarak gösterilir. Gövdenin son ünlüsü o olması durumunda, AQ yüzeyde oq olarak gösterilmektedir.

AQ:aq => V:aC*+C*__

AQ:oq => V:oC*+C*__

tuta+AQ	V(alevlen)+VVD_AQ
tut00aq	
tutaq	çok sinirlenmek
bosh+AQ	N(baş)+NND_AQ
bosh0oq	
boshoq	başak

5. Özbekçede bir gövde q harfini içeriyor ve son ünlüsü a harfi ise, I ={u, i} içeren bir ek alması durumunda son ünlü a harfi yüzeyde o harfine dönüşür. Benzer şekilde bir gövde q harfini içeriyor ve gövdenin son ünlüsü i harfi ise, bu durumda da I meta harfi yüzeyde u harfine dönüşür.

a:o => C:q__+IC
i:u => C:q__+IC:q

qara+Iv	V(bakmak)+INF_IV
qaro00v	
qarov	bakış
oqi+Iv	V(okumak)+ INF_IV
oqu00v	
oquv	okuma

6. C_s sert sessizlerinden biri ile biten gövdeye D={d, t} ile başlayan bir ek geldiğinde D meta harfi yüzeyde t sessizi ile gösterilir. C_s sert sessizlerinden biri ile bitmeyen gövdelere eklenen ekin başındaki D meta harfi d sessizi ile gösterilir.

D:t <=> C_s+__
D:d <=> C_y+__

ayt+Dir	V(söyle)+CAUSDIR_DIR
ayt0tir	
ayttir	birine söyletmek
qil+Dir	V(yap)+ CAUSDIR_DIR
qil0dir	
qildir	birinin yapmasını sağlattırmak

7. Son ünlüsü o veya u olan gövdelere Q={q, k} meta harfini içeren bir ek gelirse, Q meta harfi yüzeyde q olarak gösterilir. Gövdenin son ünlüsü a, e veya i harflerinden biri ise ve Q meta harfini içeren bir ek alırsa Q meta harfi yüzeyde k olarak gösterilir. Bu kuralın istisnaları bulunmaktadır. Örneğin, chush kelimesi Qir ekini aldığında yüzeyde chushkir olarak görünür. Bu örnekte Q meta harfi q olması gerekirken yüzeyde k olarak görünmektedir.

Q:q => V:[o | u]C*+C?__C*
Q:k => V:[a | e | i] C*+C?__C*

o'kir+IQ	V(bağır)+NVD_IQ
o'kir0ik	
o'kirik	ağlama
solin+chAQ	V(kon)+JVD_CHAQ
solin0chak	
solinchak	kulak memesi
chush+Qir	N(fış)+VND_QIR
chush0kir	
chushkir	hapşır

8. Özbekçede, k veya g ünsüzleri ile biten gövdelere G={q, k, g, g'} meta harfiyle başlayan bir ek gelirse, G meta harfi yüzeyde k olarak gösterilir. Gövdenin son ünsüzü g' veya

C_s sert sessizlerinden biri ise, ekin başında bulunan G meta harfi yüzeyde q olarak gösterilir. Ayrıca e, k veya g harflerinden birini barındıran gövdeye G meta harfiyle başlayan bir ek gelirse G meta harfi yüzeyde g olarak gösterilir. Ek olarak, o, q, x veya g’ harflerinden birini barındıran gövdeye G meta harfiyle başlayan bir ek gelirse G meta harfi yüzeyde g’ olarak gösterilir.

G:q => @[g’ | C_s]+__
G:k => @[k | g]+__
G:g => C*V*[e | k | g]V*C*+__
G:g’ => C*V*[o | q | x | g’]V*C*+__

sot+Gin	V(satmak)+NVD_GIN
sot0qin	
sotqin	hain
turt+Gi	V(dürt)+NVD_GI
turt0ki	
turtki	dürtü
yul+Gich	V(yol)+NVD_GICH
yul0g’ich	
yulg'ich	hırsız
ayt+Gali	V(söylemek)+AVD_GALI
ayt0gali	
aytgali	söylemek için

9. Özbekçede istisnai bir durum olarak bazı gövdelere sesli harfle başlayan bir ek geldiğinde gövdenin sonundaki k ve q harfleri yüzeyde sırasıyla g ve g’ harfine dönüşür.

k:g => __+V
q:g’ => __+V

terak+Am	N(kavak ağacı)+NNI_POSS1S_AM
terag0im	
teragim	kavak ağacım
quloq+imiz	N(kulak)+NNI_POSS1P_AMIZ
qulog’0imiz	
qulog'imiz	kulağımız

10. Ünlü ile biten gövdelere ünlü ile başlayan bir ek geldiğinde, ekin başındaki ünlü yüzeyde düşer.

V:0 => V+__

bola+Am	N(çocuk)+NNI_POSS1S_AM
bola00m	
bolam	çocuğum
o'qi+ib+man	V(oku)+TPASTNARR_IB+PERS1S1_MAN
o'qi00b0man	
o'qibman	okumuşum

11. Sessiz ile biten bir gövdeye n sessizi ile başlayan bir ek geldiğinde, ekin başındaki n ünsüzü yüzeyde düşer.

n:0 <=> C+__

men+ning	R(ben)+NNI_GEN_NING
men00ing	
mening	benim
sen+ni	R(sen)+NNI_OBJ_NI
sen00i	
seni	seni

12. Bir istisna kural, yirgan ve irgan kelimeleri için mevcuttur. Buna göre bu kelimelere fiilden sıfat yapan yapım eki -chiQ eklendiğinde kelimelerin sonundaki n sessizi m sessizine dönüşür.

n:m <=> C*V*__+chiQ

irgan+chiQ	V(iğren)+JVD_CHIQ
irgam0chik	
irgamchik	iğrenç
yirgan+chiQ	V(iğren)+JVD_CHIQ
yirgam0chik	
yirgamchik	iğrenç

13. Bazı kelimelerde, gövdenin son ünlüsü, ünlüyle başlayan bir ek aldığında yüzeyde düşer. Bazı durumlarda, ünsüzle başlayan bir ek alan gövdenin ünsüz harfi düşer.

V:0 => __C*+V
C:0 => __+C*

burun+Am	N(burun)+NNI_POSS1S_AM
bur0n0im	
burnim	burnum
suv+GAr	N(su)+VND_GAR
su00g'or	
sug'or	sulamak

14. Özbekçede özel bir durum olarak, bir sayısı -ta sıfat yapan yapım ekini aldığında, sayının sonundaki r sessizi t sessizine dönüşür.

bir+ta	N(bir)+JND_TA
bit0ta	
bitta	bir tane

15. Özbekçede fiilden sıfat yapan -chAQ eki ile ilgili özel bir durum söz konusudur. Buna göre ekte yer alan A meta harfi yüzeyde o veya a olarak gösterilir. Ekin eklendiği gövdenin ilk seslisi a, o, i veya u seslilerinden biriyse ek -chok olarak, diğer seslilerin olması durumlarda -chak olarak gösterilir.

ilin+chAQ	V(takıl)+JVD_CHAQ
ilin0choq	
ilinchoq	tutan şey
em+chAQ	V(em)+ JVD_CHAQ
em0chak	
emchak	göğüs

16. Özbekçede, g veya g’ ünsüzleri ile biten gövdelere G={q, k, g, g’} meta harfiyle başlayan bir ek gelirse, G meta harfi yüzeyde sırasıyla k veya q olarak gösterilir. Bu kurala istisna olarak Rusçadan geçen kelimeler gösterilebilir, biolog kelimesi bu kurala istisna olarak aşağıda sunulmuştur.

g:k => __+G
g’:q => __+G

teg+Gani	V(değmek)+AVD_GANI
tek0kani	
tekkani	değmek için
bog'+Ga	N(bahçe)+NNI_DAT_GA
boq0qa	
boqqa	bahçeye

biolog+Ga	N(biyolog)+ NNI_DAT_GA
biolog0ga	
biologga	biyoloğa

17. Özbekçede iki ünlü yan yana gelmez. Bazı durumlarda, ünlü ile biten gövdelere ünlü ile başlayan bir ek geldiğinde y ünsüzü iki ünlü arasına eklenir.

parvo+Imiz	N(ilgi)+NNI_POSS1S_IM
parvo0yimiz	
parvoyimiz	ilgimiz

mavqe+Ing	N(mevki)+ NNI_POSS1S_IM
mavge0ying	
mavqeying	senin mevkin

18. Bu kurala göre, Y={a, y} meta harfi ile başlayan ek sessizle biten bir gövdeye gelirse Y yüzeyde a olarak gösterilir. Bu ek sesli ile biten bir gövdeye gelirse Y yüzeyde y olarak gösterilir.

Y:a => C+__
Y:y => V+__

kel+Ydigan	V(gel)+JVD_YDIGAN
kel0adigan	
keladigan	gelecek

arzi+Ydigan	V(kıymetli ol)+ JVD_YDIGAN
arzi0ydigan	
arziydigan	kıymetli

19. Ünsüzle biten gövdelere ünsüzle başlayan bir ek geldiğinde ekin başındaki ünsüz yüzeyde düşer.

C:0 => C+__

bosh+si	N(baş)+NNI_POSS3S_SI
bosh00i	
boshi	onun başı

kitob+viy	N(kitap)+ JND_VIY
kitob00iy	
kitobiy	onun kitabı

20. Qizil sıfat gövdesinin son hecesi(-il), gövdeye -ar, -r veya g’ish eklerinin eklenmesi ile düşer. Bu durum yüzeyde gösterilir.

qizil+g’ish	J(kırmızı)+JJD_GISH
qiz000g’ish	
qizg'ish	kırmızımsı

qizil+ar	J(kırmızı)+VND_AR
qiz000ar	
qizar	kızarmak

21. Kichik sıfat gövdesinin son hecesi (-ik), gövdeye -gina, -kina, -qina, -g’ina veya g’ish eklerinin eklenmesi ile düşer. Bu durum yüzeyde gösterilir.

kichik+Gina	J(küçük)+JJD_GINA
kich000kina	
kichkina	küçücük

22. Qati sıfat gövdesine -viy ve -iy ekleri eklenince, eklerin başındaki harfler düşerek bu ekler yüzeyde -y olarak gösterilir.

qati+viy J(kati)+JJD_(V)IY
qati000y
qatiy kesinlikle

23. Chinakam ve rostakam sıfat gövdelerine -asiga, -isiga, ve -siga ekleri eklenince, eklerin başındaki harfler (as, is ve i) düşerek yüzeyde gösterilir.

chinakam+asiga J(gerçekten)+AJD_(A/S)IGA
chinakam000iga
chinakamiga gerçekten de

rostakam+asiga J(hakikaten)+ AJD_(A/S)IGA
rostakam000iga
rostakamiga hakikaten de

24. U (o), bu (bu), shu (şu) ve o'sha (o) zamirlerine iyelik ekleri eklenirse, zamir ile ek arasında ni sesi türer. Benzer şekilde, bu zamirlere eklenen diğer eklerde de zamir ile ek arasına n sesi türer.

shu+im R(şu)+NNI_POSS1S_IM
shu0ni0m
shunim şuyum

o'sha+si R(o)+ NNI_POSS3S_IM
o'sha0nisi
o'shanisi işte onunkisi

bu+da R(bu)+NNI_LOC_DA
bu0nda
bunda bunda

shu+gacha	R(şu)+ADJ_GACHA
shu0ngacha	
shungacha	şuna kadar

25. Özbekçeye yabancı dillerden geçen ve kelimenin son iki harfi l olan gövdelere, l ile başlayan bir ek geldiğinde, gövdenin son l harfi düşer.

l:0 <=> l__+lV*C*

xlorofill+lar	N(klorofil)+NNI_PLU_LAR
xlorofil00lar	
xlorofillar	klorofiller
metall+lar	N(metal)+NNI_PLU_LAR
metal00lar	
metallar	metaller

26. Men ve sen zamirlerine -ning, -ni ve -niki ekleri eklendiğinde, zamirlerin sonlarındaki n harfi yüzeyde silinir.

n:0 <=> n__+nV*C*

men+niki	R(ben)+NNI_REL3_NIKI
me00niki	
meniki	benim
metall+lar	N(metal)+NNI_PLU_LAR
metal00lar	
metallar	metaller

3.1.2. Özbekçenin Biçim Dizgesi

Biçim dizgesi (morphotactics), doğal bir dilde biçim birimlerin (morfemlerin) düzeni, sıralaması ve kısıtlamaları tanımlar (Sproat, 1992). Özbekçede biçim birimler kök

kelimeye/gövdeye ve birbirlerine belirli düzen ve sıra içerisinde eklenir. Özbekçenin biçim birimleri ve sıralaması incelendiğinde bazı istisnalara rastlanmakta olup, bu bölümde biçim dizgesi ile istisnaları da ayrıca anlatılmaktadır.

Sonlu durum makineleri (FSM), biçim dizgesinin gösterimi için kullanılan ana modellerden biridir. Koskenniemi (1984) iki düzeyli modeli, biçim birimlerin sıralanmasını ve düzenini sonlu durum makineleri ile tanımlar. Sonlu durum makineleri durumlar (state) ve bu durumlar arasında geçişlerden (transitions) oluşmaktadır.

Özbekçe biçim dizgesi, bu modele uygun olarak bir dizi kurallar ile sonlu durum makineleri ile sunulmuştur Buna göre, başlangıç durumu (initial state) isim (nominal) ve fiil (verbal) kökü olabilir. İsim (nominal) kök, isimler, sıfatlar, zarflar vb. sözcük türlerinden oluşur. Fiil (verbal) kök, fiil kökenli köklerden veya isim (nominal) köklerden türetilen fiil köklerden oluşur. Kelimelerin, sözcük türleri (POS) etiketleri durumlar (states) olarak sunulmaktadır. Yönlü bir graf olan sonlu durum makinelerinde sonraki duruma, o duruma eklenen biçim birimler (morfemler) aracılığı ile yönlendirilerek (direct edges) bir yönde (ok yönünde) geçilir. Durumlar (states) arasındaki geçişlerde, geçiş yönündeki ok biçim birimin (morfemin) sözcüksel(lexical) biçimi ile etiketlenir. Bazı durumlarda iki durum arasındaki geçişlerde yönlü ok 0 ile etiketlenir, bu sonraki duruma geçmek için bir ek(biçimbirim) kullanılmadığını gösterir ve sonraki duruma direk geçilir. Başlangıç durumundan son duruma kadar yönlü oklar ile geçilen tüm durumlar ara durumlar (intermediate states) olarak ifade edilir. Son durum (final state) hiçbir ek(biçimbirim) ile etiketlenmeyen sonuncu durum olup, gövdenin son halini göstermektedir.

Bu bölümün alt bölümlerinde isim (nominal) ve fiil (verbal) biçim dizgesi kurallarla, örneklerle ve sonlu durum makineleri ile detaylandırılmaktadır. Özbekçenin biçim dizgesi araştırılırken gramer kurallarına ve paradigmasına uymayan birçok özel durum ve istisna ile karşılaşılmış olup bu durumlar sonraki alt bölümlerde sunulmaktadır. Ayrıca Özbekçenin biçim dizgesi tanımlamaları uygulama aşamasında kullanılmakta olup, biçim bilimsel çözümleyicimiz Nuve Studio kök kelime ve ekler sözlüğünün yanında, biçim dizgesi kurallarını, düzenini ve sıralamasının da tanımlanması, uygulama gereksinimleri arasında yer almaktadır. Ayrıca, Nuve Studio biçimbilimsel çözümleme için biçim dizgesi kurallarının yanında yazım (ortografik) kuralların tanımlanmasını da morfolojik üretim için kullanmakta olup uygulama ile ilgili detaylar sonraki bölümlerde sunulmaktadır. Bu bölümde Özbekçenin biçim bilimsel

çözümlemesi için tanımlaması yapılan ve detaylandırılan her bir konu, kural, paradigma ve gramer yapısı uygulama aşamasında kullanılmıştır. Biçim dizgesi ve sonlu durum makineleri de uygulama aşamasında birebir kullanılarak Özbekçenin Nuve Studio'da yer alan morphotactics (biçim dizgeleri) dosyası oluşturulmuş ve bu sayede biçimbilimsel çözümlenmesi sağlanmıştır.

3.1.2.1. İsim Köklü Biçim Dizgesi

Bu bölümde Özbekçenin, isim (nominal) soylu kelimelerin kök olduğu, biçim dizgesi detaylandırılarak, örnekler, istisnalar ve isim köklü biçim dizgesinin kuralları model(paradigma) olarak sunulmuştur. Özbekçenin isim soylu (nominal) paradigması sonlu durum makineleri ile sunulmuştur. İsim soylu (nominal) kök, isimler, sıfatlar, zarflar vb. sözcük türlerinden oluşur.

Özbekçenin isim köklü biçim dizgesi düzenli kurallarına (nominal paradigma) göre, isim köklerine-gövdesine eklenen çekim ekleri düzenli ve sıralı olarak eklenmektedir. Bu paradigma köke-gövdeye gelen eklerin sırasını düzenlediği gibi eklerin birbirleriyle olan sırasını da düzenlemektedir. Bu durumun birçok istisnası olmasına rağmen isim kökü sırasıyla, çoğul, iyilik, hâl eki, ilgi ve soru ekleri düzeninde ekler almaktadır. Özbekçede, temel olarak aşağıdaki ek sırası düzenli kuralına uyulması beklenir.

İsim + Çoğul Eki + İyelik Eki + Hâl Eki + İlgi Eki + Soru Eki

Özbekçede, istisnalar olsa da çoğu kelime bu paradigmaya uygun bir şekilde oluşturulur. Paradigmada gösterilen model için ilgi eki ve hâl eki ile ilgili bir detay bulunmaktadır. İlgi eki hâl eklerinden sadece -da bulunma hâl ekinden sonra gelir, diğer hâl eklerinden sonra gelmez. Kitoblarimda (kitaplarımda) kelimesi ve eklerin sıralanışı bu paradigmaya uygun bir örnektir. Örneğin biçim bilimsel çözümlemesi aşağıdaki gibidir.

kitob (isim) + lar (çoğul eki) + im (1. tekil kişi iyelik eki) + da (bulunma hâl eki)

Bir çok kelime bu paradigmaya uygun olarak ekler almaktadır. Özbekçe gramerindeki düzensiz durumlar ve istisnalar Özbekçe biçim dizgesinde de karşımıza çıkmaktadır. Aşağıda Özbekçe isim (nominal) köklü biçim dizgesinde yer alan istisna ve özel durumlar açıklamaları ve örneklerle sunulmuştur.

- Özbekçede çoğul eki, iyelik ekinden sonra gelmez. Çoğul ekinin, iyelik ekinden sonra geldiği durumlar istisna durumlar olup, isim köklü biçim dizgesi paradigmasına uymaz.

dadamlar (babamlar)
dada (isim) + m (1.tekil kişi iyelik eki) + lar (çoğul eki)

- İlgi ekleri sadece -da bulunma hâl ekinden sonra eklenir, diğer hâl eklerinden sonra eklenmez. Bu özel durum aşağıdaki örnekte gösterilmektedir.

uydagi (evdeki)
uy (isim) + da (bulunma hâl eki) + gi(ilgi eki)

- İsim (nominal) köklü paradigmaya göre ilgi eki, çoğul ekinden sonra gelmelidir. Bazı istisna durumlarda çoğul eki, ilgi ekinden sonra gelebilir. İlgi eki -gi için istisna bir durum söz konusudur. Aşağıdaki örnekte bu durumlar gösterilmektedir.

kitobimdagilar (kitabımdakiler)
kitob (isim) + im (1. tekil kişi iyelik eki) + da (bulunma hâl eki) + gi (ilgi eki) + lar (çoğul eki)

Bu örnekte birden fazla istisna bulunmaktadır. İstisnalardan ilki çoğul eki, iyelik ekinden sonra eklenmiştir. İkinci istisna çoğul eki hâl ekinden sonra eklenmiştir. Son olarak çoğul eki, ilgi ekinden sonra eklenmiştir. Bu örnek birçok istisnanın yanında birde özel durum içermektedir. Bu özel durum ilgi eki sadece -da bulunma hâl ekinden sonra eklenir, diğer hâl eklerinden sonra eklenmez, özel durumudur.

- Birinci tekil şahıs eki hâl eklerinden sonra gelmez. Bu kurala istisna durum aşağıdaki örnekte gösterilmiştir.

shahardanman(şehirdenim)
shahar (isim) + dan (ayrılma hâl eki) + man (1. tekil kişi)

Özbekçede birleşik isimler genel olarak ayrı yazılmaktadır. Birleşik isimlerde İsim (nominal) köklü paradigmaya uygun olarak ek alır. Özbekçede birleşik isimlerde ek ikinci isme eklenmektedir. Birleşik isimler için örnekler aşağıda gösterilmiştir.

cheoy qoshg'i (çay kaşığı)

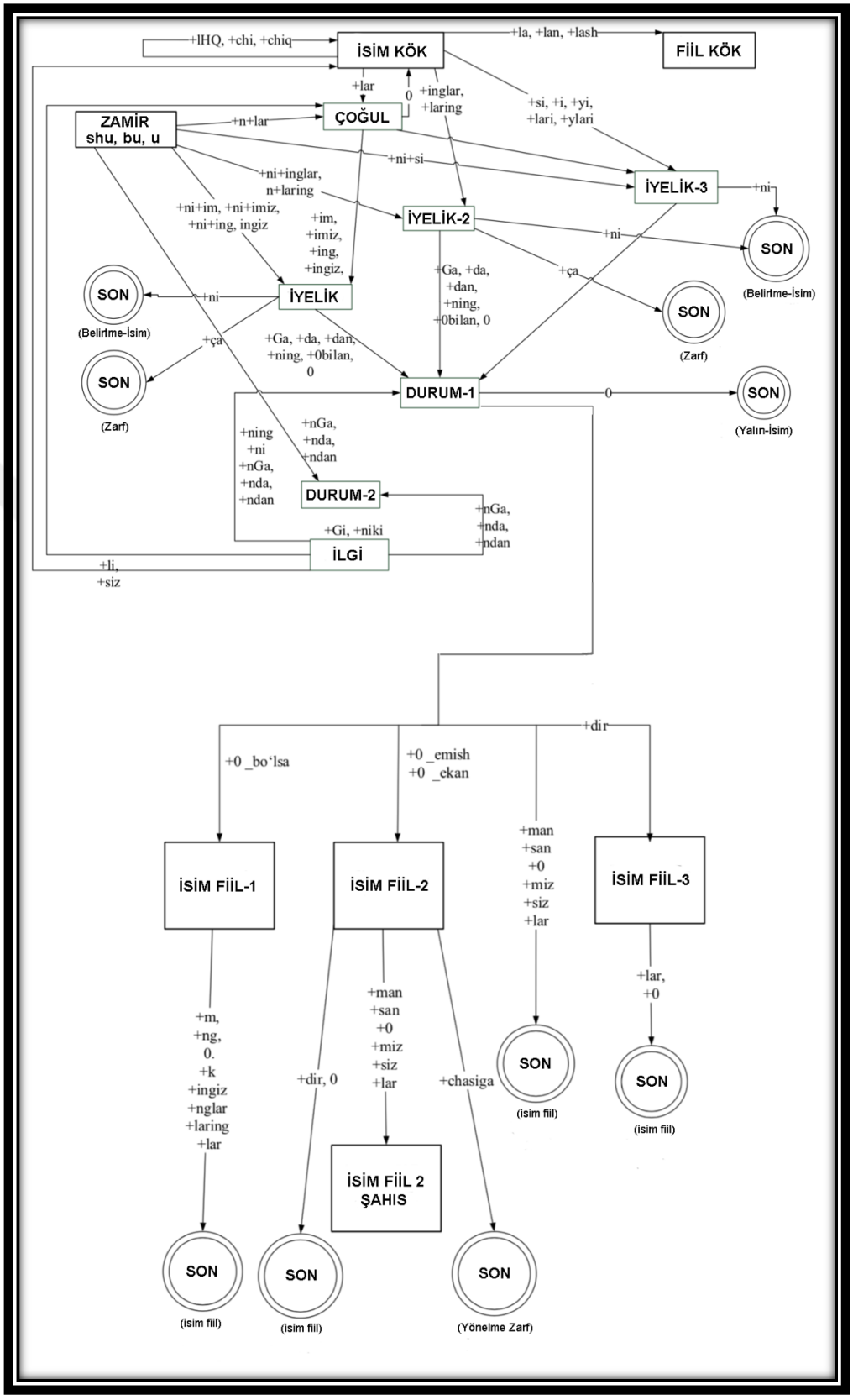
cheoy (isim) + qoshig'(isim) + i (belirtme hâl eki)

O'zbek eding (sen Özbektin)

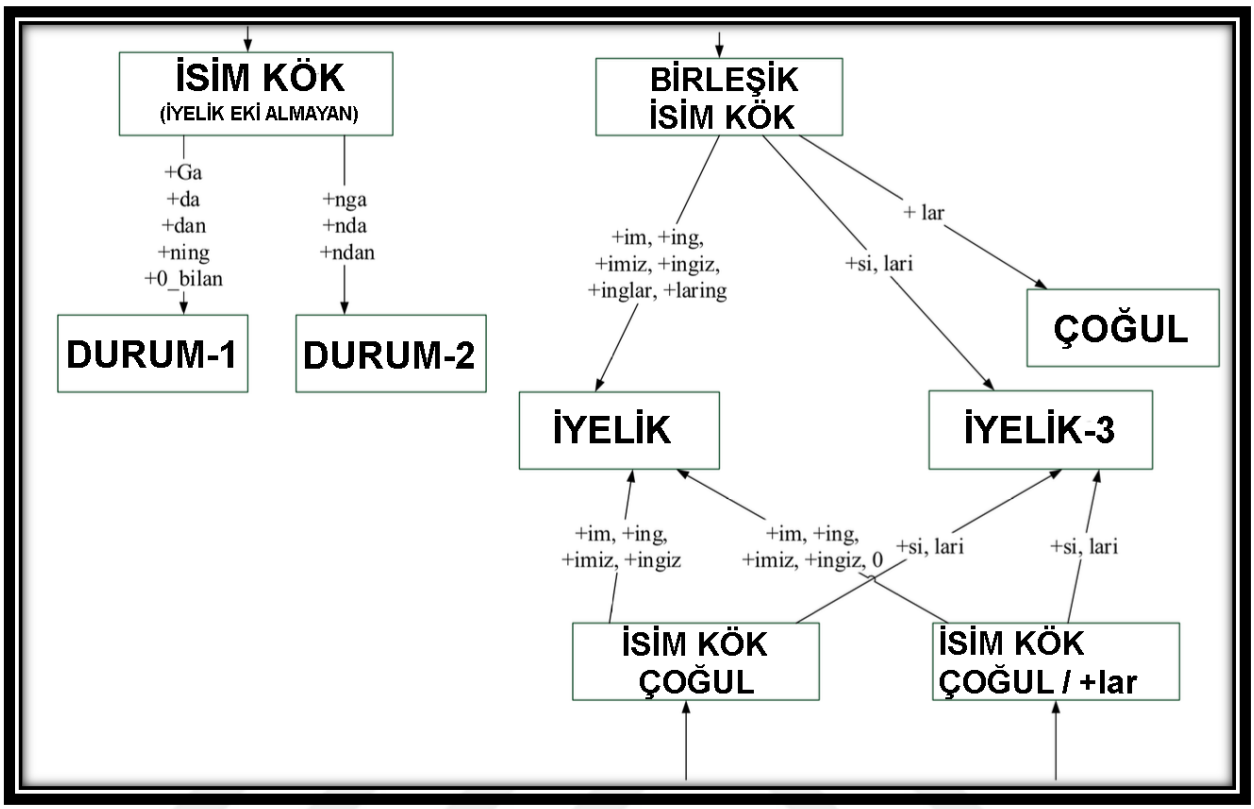
O'zbek (isim) + ed(isim) + ing(iyelik eki)

Özbekçenin isim (nominal) kök biçim dizgesini gösteren sonlu durum makinesi Şekil 3.1'de gösterilmiştir. Birleşik isimleri ve onların biçim dizgesini ait sonlu durum makinesi Şekil 3.2'de sunulmaktadır. Sonlu durum makinelerinde gösterildiği üzere, her bir durumdan (state) bir başka duruma eklenen ekler ile geçilmekte olup, bu eklerin eklendikten sonra geçilen durum yönlü oklar ile gösterilmiştir. Her bir yönlü ok üzerinde bulunan etiketler ekleri göstermektedir. Kök (başlangıç) durumuna çekim ekleri (morfemler) paradigmaya uygun olarak eklenerek son duruma (final state) kadar geçiş sağlanmaktadır. Buna göre çekim ekleri sırasıyla, çoğul eki(-lar), iyelik eki (-im, -ing vb.), hâl eki (-ni, -ning vb.), ilgi eki (-ki, -niki vb.) ve soru eki (-mi) sırasıyla eklenerek başlangıç durumundan, son duruma geçiş sağlanmaktadır. Türetme (yapım) ekleri eklendikleri gövdeden yeni gövde türeten eklerdir. İsimden fiil (-la, -lan vb.), isimden zarf (-lay, -lab vb.), isimden sıfat (-li, -siz vb.), isimden isim (-chi, -chiq vb.), sıfattan isim (-cha, -chik vb.), sıfattan zarf (-asiga, -kan vb.), sıfattan sıfat (-imtir, -ish vb.) ve zarftan zarf (-day, -dak vb.) ekleri türetme ekleridir.

Sonraki alt bölümde, Özbekçenin fiil(verbal) köklü biçim dizgesi kurallarla (paradigma), örneklerle ve sonlu durum makineleri ile detaylandırılmaktadır. Bu bölümde anlatılan kurallar ve sonlu durum makineleri uygulama aşamasında Nuve Studio morphotactics dosyasının oluşturulmasında kullanılmış olup, sonraki bölümlerde uygulama detayları ayrıca sunulacaktır.



Şekil 3.1: İsim kökleri için sonlu durum makinesi (FSM) modeli.



Şekil 3.2: Birleşik isimlerin biçim dizgesi için sonlu durum makinesi.

3.1.2.2. Fiil Köklü Biçim Dizgesi

Bu bölümde Özbekçenin fiil (verbal) köklü biçim dizgesi kurallarla (verbal paradigma), örneklerle ve sonlu durum makineleri ile sunulmuştur. Özbekçede fiil köklü biçim dizgesi fiil tabanlı gövdelerden ve fiil olarak türetilen (isimden türetilen fiil, sıfattan türetilen fiil vb.) gövdelerden meydana gelmektedir.

Özbekçede fiil (verbal) köklerine-gövdelerine eklenen ekler belirli bir sıra ve düzenle eklenmektedir. Bazı istisnalar olmasına rağmen genel olarak fiil köküne-gövdesine eklenen eklerin belirli bir sıra ve düzeninin yanında, eklerinde birbirleri ile arasındaki ilişkinin bir düzeni vardır. Bu düzen fiil köklü biçim dizgesinin paradigmasını oluşturur. Özbekçede fiil köküne fiil yapım eki, ardından fiil çekim eki gelir. Fiil yapım (türetme) eki yoksa, fiil çekim eki doğrudan fiil köküne eklenebilir. Fiil çekim ekleri genellikle fiil köküne belirli bir sıra ile eklenir, bu sıra fiil köklü biçim dizgesinin paradigmasını oluşturmakta olup aşağıdaki gibidir.

Fiil + Fiil Çatı Eki + Olumsuzluk Eki + Zaman Eki + Şahıs Eki

Özbekçenin fiil köklü biçim dizgesinin paradigması, fiil köküne sırasıyla gelen fiil çatı, olumsuzluk, zaman ve şahıs ekinden oluşmaktadır. Fiil çatı ekleri (etken, edilgen, ettirgen vb.), olumsuzluk eki (-ma), zaman ekleri (şimdiki, gelecek, geniş, gelecek-geniş, görülen geçmiş, yakın geçmiş zaman, öğrenilen geçmiş zaman, gereklilik, şart) ve şahıs ekleri (tekil ve çoğul şahıs ekleri; -man, -miz, -san, -siz, -sizler, -laring, -di, -ayapti, -dilar, -m, -k, -ng, -ngiz, - ngizler, -nglar, -laringiz, -ish ve -lar) sırasıyla fiil köküne ve ekler her birine eklenmektedir. İkinci çoğul şahıs eklerinden -siz eki resmi olmayan ve -sizlar eki resmi olarak kullanımını göstermektedir.

Özbekçe fiil köklü biçim dizgesinin paradigmasına uygun olarak eklerin sıralandığı keltirittirilmagansan (getirttirilmemişsin) ve ayrı olarak yazılan qaratgan ekanman (baktırmışım) örnekleri biçimbilimsel olarak çözümlenerek aşağıda sunulmuştur.

keltirittirilmagansan (getirttirilmemişsin)

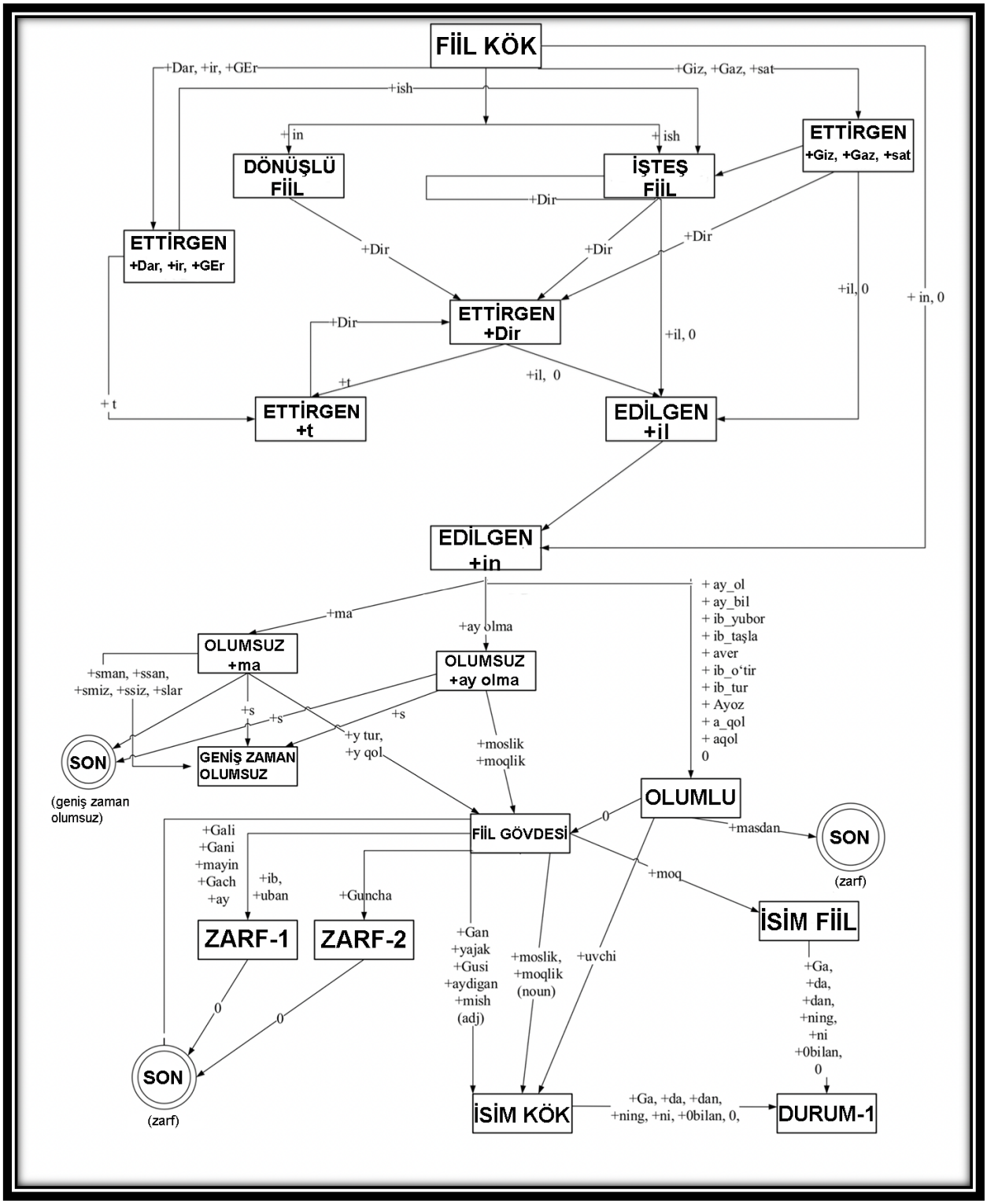
kel(fiil) + tir(ettirgen fiil) + it(ettirgen fiil) + tir(ettirgen fiil) + il(edilgen fiil) + ma(olumsuz) +gan(geçmiş zaman) + san (2. tekil kişi)

qaratgan ekanman (baktırmışım)

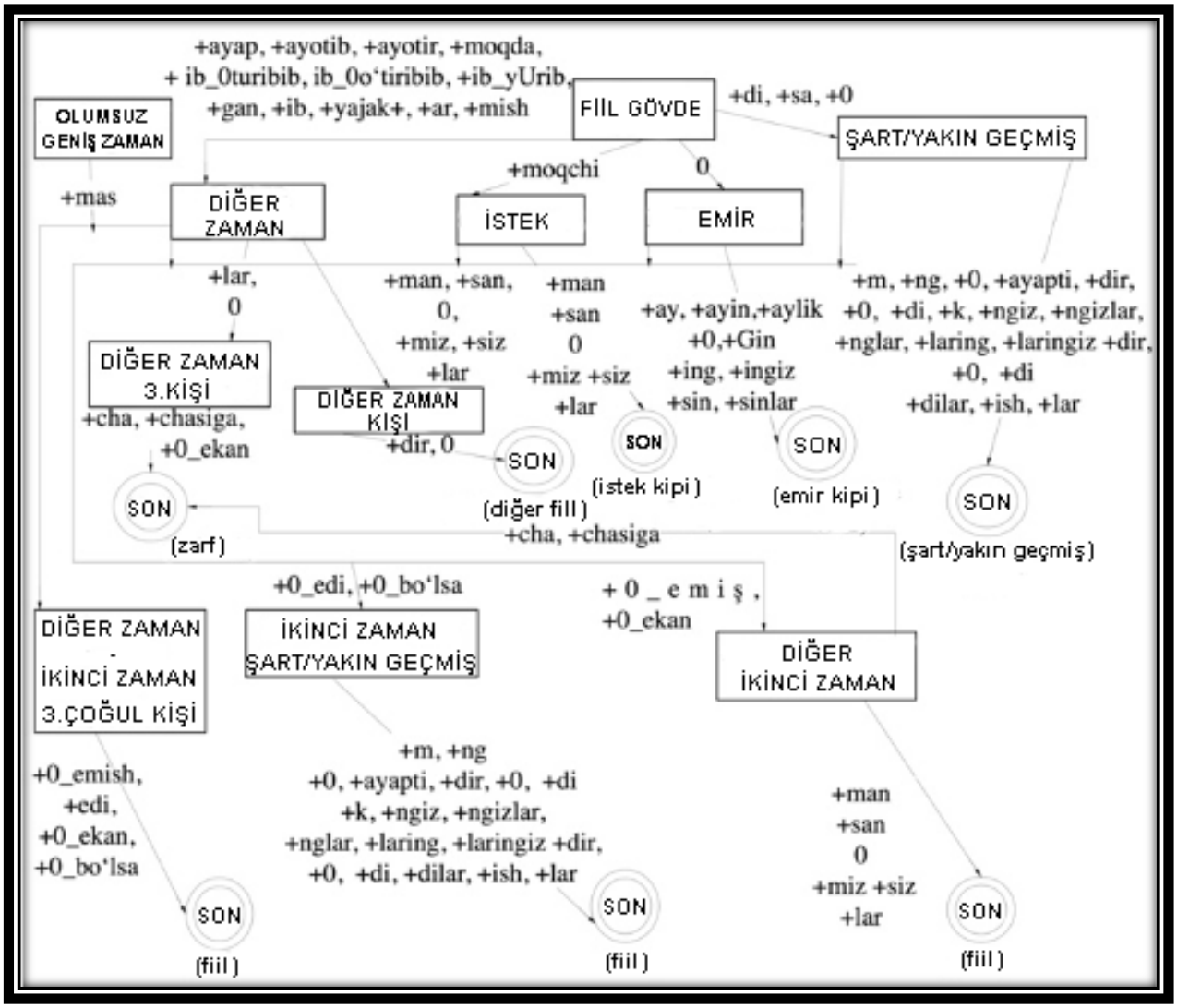
qara(fiil) + t(ettirgen fiil) + gan(geçmiş zaman) + ekan (rivayet geçmiş zaman) + man (1. tekil kişi)

Özbekçenin fiil köklü biçim dizgesi Şekil 3.3'te sonlu durum makineleri ile gösterilmektedir. Başlangıç (kök) durumundan diğer durumlara ok yönündeki geçişler, ok üzerinde etiketlenen ekler ile sağlanmaktadır. Fiil çekim ekleri, fiil köklü biçim dizgesi paradigmasında gösterildiği üzere sırasıyla, çatı ekleri (-ir, -sat vb.), olumsuzluk eki (-ma), zaman ekleri(-di, -gan vb.) ve şahıs ekleri (-man, -san vb.) fiil köküne eklenmektedir. Türetme(yapım) ekleri yeni gövdeler türetmektedir. Fiilden isim (-uv, -inch vb.), fiilden sıfat (-ar, -asi vb.), fiilden fiil (-ay, -imsira vb.) ve fiilden zarf (-ib, -uban vb.) türetme ekleridir.

Özbekçenin isim ve fiil köklü biçim dizgesi ve detayları örnekler, sonlu durum makineleri, kurallar ve istisnalar ile açıklanmıştır. Sonraki bölümde Özbekçenin biçimbilimsel çözümlemesinde ortaya çıkan belirsiz durumlar üzerinde durularak, belirsizlik giderme yöntemine ait detaylar paylaşılacaktır.



Şekil 3.3: Fiil kökleri için sonlu durum makinesi (FSM) modeli.



Şekil 3.3 (devam):

3.2. ÖZBEKÇE İÇİN BİÇİMBİLİMSEL BELİRSİZLİK GİDERME

Tezin bu bölümünde Özbekçenin biçimbilimsel belirsizlik gidermesi için kullanılan yöntem ve detayları paylaşılmıştır. Uygulama ile ilgili hazırlık detayları, uygulama aşamaları, çıktıları ve elde edilen bulgular sonraki bölümde sunulmaktadır. Özbekçenin biçimbilimsel çözümlemesi sonrası, biçimbilimsel belirsizlik giderme ihtiyacı ortaya çıkmıştır.

Özbekçenin biçimbilimsel çözümlemesi bölümünde ve uygulamasında karşılaşıldığı üzere bir kelimenin birden fazla biçimbilimsel çözümlemesi bulunmaktadır. Nuve Studio mümkün olan tüm çözümleri getirebilecek şekilde programlanmıştır. Ancak bu çözümlerden hangisi ilgili metin ya da durum için en doğru ve istenilen çözüm olduğunun seçilmesi biçimbilimsel belirsizlik gidermenin konusudur.

Bu bölümde Özbekçenin belirsizlik giderme probleminin tanımı ve belirsizlik giderme için kullanılan yöntem alt bölümlerde detayları ile paylaşılmıştır. Detayları paylaşılan yöntemin uygulaması ve uygulama çıktıları sonraki bölümde anlatılmaktadır.

3.2.1. Özbekçe için Biçimbilimsel Belirsizlik Giderme Probleminin Tanımı

Biçimbilimsel belirsizlik giderme, bir kelimenin (bir metindeki veya yalın), birden fazla biçimbilimsel çözümü olması durumunda ilgili kelimenin biçimbilimsel çözümünün tek bir çözüme istenilen (doğru) çözüme indirgenmesi ve tek bir çözümün çıktı olarak gösterilmesini sağlayan metotlardır. Aşağıdaki örnekte o'rma kelimesinin Nuve Studio biçimbilimsel çözümleyicisi ile üretilen iki ayrı biçimbilimsel çözümü gösterilmektedir.

Kelime	**Biçimbilimsel Çözüm**	**Açıklama**
o'rma	FİİL(o'r)+FY_ISIM_mä(ma)	Örgü anlamına gelmektedir. Örmek fiiline fiilden isim yapan -ma eki getirilmiştir.
o'rma	FİİL(o'r)+OLUMSUZLUK_ mä (ma)	Örme olumsuz anlamına gelmektedir. Örmek fiiline olumsuzluk eki – ma getirilmiştir.

Yukarıdaki örnekte gösterildiği üzere yazılışı aynı olan kelimenin iki farklı biçimbilimsel çözümü bulunmaktadır. Örnekteki iki biçimbilimsel çözümde doğru olarak çözümlenmiştir. Nuve Studio biçimbilimsel çözümleyicisi tarafından üretilen bu çözümlerden bir tanesi seçilmek istendiğinde biçimbilimsel belirsizlik metotlarının uygulanması gerekmektedir.

Birden fazla biçim bilimsel çözümden istenilen, seçilen veya en doğru olan çözümü bulmak için uygulanacak biçimbilimsel çözümleme metotlarının girdi olarak kabul ettiği veriler bir biçimbilimsel çözümleyici tarafından çözümlenmiş farklı biçimbilimsel çözümlerdir.

Yukarıdaki örnekte iki farklı çözümlemeden hangisinin en doğru çözüm olduğunu bulmak için bir cümle içerisinde kullanımlarının sunulması ile tam olarak hangi çözümün istenilen ya da en doğru çözüm olduğunu anlayabilmekteyiz. Bir olumsuzluk anlamı olan kelimenin ve fiilden isme türetilen bir kelimenin önceki ve sonraki kelimelerinden ve onların çözümlerinden

istenilen çözüm anlaşılabilir. Daha karmaşık problemler için daha farklı yöntemler kullanılabilir.

Bu çalışmada Dayanık ve diğ. (2018) çalışması referans alınarak Özbekçe için biçimbilimsel belirsizlik giderme sağlanmıştır. Detayları sonraki alt bölümde anlatılan MorphNet ile hem biçimbilimsel çözümleme hem de biçimbilimsel belirsizlik gidermenin tek bir model ile çözümlenebilirken, çalışma kapsamında sadece belirsizlik giderme için bu model kullanılmıştır.

3.2.2. Özbekçe için Biçimbilimsel Belirsizlik Giderme ve Kullanılan Yöntem

Özbekçenin biçimbilimsel belirsizlik gidermesi için MorphNet (Dayanık ve diğ., 2018) kullanılarak, Nuve Studio ile biçimbilimsel çözümlemesi yapılan Özbekçe kelimelerin, biçimbilimsel çözümlemesi MorphNet ile sağlanmıştır.

Yapay sinir ağlarının belirsizlik giderme için kullanıldığı yöntemlere göre kelime kök ve morfemleri öğretilerek sonrasında test için verilen bir kelimenin önceki ve sonraki kelimelerini de kontrol ederek doğru çözümün tahmini sağlanmaktadır. Tolue ve diğ. (2017), Türkçe ve Kazakça için yapay sinir ağları modeli kullanarak biçimbilimsel belirsizlik giderme sağlamışlardır. Evrişimsel sinir ağları (CNN) ve LSTM (Long Short Term Memory) tabanlı tekrarlı sinir ağları kullanılarak doğru çözüm tahmini yapılmaktadır.

Bu çalışmada MorphNet modeli referans alınmıştır. MorphNet girdi olarak aldığı metnin içerisindeki kelimelerin doğru biçimbilimsel çözümünü üreten, bir sıralı (seq2seq) tekrarlayan yapay sinir ağı (RNN) modelidir. MorphNet girdi olarak aldığı metinin içerisinde biçimbilimsel çözümleme ve sonrasında belirsizlik giderme işlemlerini tamamlayarak bir kelime için tek bir biçimbilimsel çözümlemeyi çıktı olarak üretmektedir. Her ne kadar referans çalışmada biçim bilimsel çözümleme içinde aynı model kullanılsa da Özbekçenin biçimbilimsel belirsizlik gidermesi için sağlanan biçimbilimsel çözümler Nuve Studio ile hazırlanmış ve hazırlanan ara bir program ile MorphNet için uygun formata dönüştürülmüştür.

MorphNet bir cümle içerisinde yer alan bütün kelimelerin biçimbilimsel çözümünü üretmektedir. Bu çözümlemede kelimenin kök ve biçimbilimsel özellikleri yer almaktadır. Üç farklı kodlayıcı kullanılmaktadır. Bu üç farklı kodlayıcı ayrı ayrı işlemlerin yerine getirilmesini sağlamaktadır. Birinci kodlayıcı kelime kodlayıcısıdır (word encoder). Kelime kodlayıcı karakterlerine göre her bir kelime için bir yapı oluşturur. Buna göre her biri karakter karakter

gömülü vektöründe haritalandırılır. Soldan sağa bütün karakterleri yer alır. Böylece biçimbilimsel çözümleme için her kelime kök ve morfemlerine karakter karakter ulaşılmaktadır. İkinci kodlayıcı bağlam kodlayıcıdır (context encoder). Bağlam kodlayıcı her bir kelimenin sol ve sağ komşu kelimelere dayanarak her kelime için bir yapı oluşturur. İki yönlü bir LSTM kullanarak her iki yönde de işlem yapabilmektedir. Böylece cümledeki her hedef kelime için eşsiz bir yapı oluşturur. Üçüncüsü çıktı kodlayıcısıdır (output encoder). Çıktı kodlayıcısı hedef kelimeden önceki kelimelerin biçimbilimsel özelliklerini yakalayan bir yapı oluşturur. Böylece hedef kelimenin türünün tespitini kolaylaştırmış olur. Bu kodlayıcılar çözümleyiciyi (decoder) besleyerek hedef kelimenin kök ve biçimbilimsel özelliklerini üretir.

MorphNet modelinin amacı ve motivasyonu tek bir model ile biçimbilimsel çözümlemeyi yapıp belirsizlik gidermeyi uygulayarak bir kelime için tek bir biçimbilimsel çözümlemeyi çıktı olarak üretmek ve kullanımı kolay bir model sunmaktır. Bu çalışmada Özbekçe için MorphNet'in belirsizlik giderme fonksiyonellikleri kullanılmıştır. Belirsizlik içeren metinler tercih edilerek Nuve Studio biçimbilimsel çözümleyicisinde metinler içerisindeki kelimelerin biçimbilimsel çözümlemeleri yapılmıştır. Sonrasında yarı otomatik bir biçimde en doğru çözümler işaretlenerek, MorphNet'e en doğru çözüm ve alternatif çözümler öğretilmiştir. Özbekçe için hazırlanan veri seti üzerinden eğitim (training) ve test sağlanmıştır.

Özbekçenin biçimbilimsel belirsizlik giderme çalışması için sırasıyla aşağıdaki işlemler uygulanmıştır.

- Özbekçe güncel metin, haber ve edebi eserler Nuve Studio kök kelime sözlüğüne eklenmiştir,
- Eklenen içeriklerin Nuve Studio ile biçimbilimsel çözümlemesi gerçekleştirilerek, MorphNet veri seti için ayrılmıştır,
- Nuve Studio çıktı formatından MorphNet girdi formatına çevirilmesi için bir ara program geliştirilmiştir,
- MorphNet veri seti için Nuve Studio biçimbilimsel çözümleyicisinden filtrelenerek elde edilen çözümler, MorphNet formatına ara program ile çevrilmiştir,
- MorphNet formatına çevrilen veri içerisinde en doğru çözüm elle seçilerek birinci çözüme yazılmıştır,

- MorphNet veri seti varsayılan olarak kabul ettiği, MorphNet geliştiricileri tarafından hazırlanan TRDataset formatına uygun olarak hazırlanmıştır. Bu formata göre kelimeler içerisinde geçtikleri cümleler içerisinde veri setine tanımlanmaktadır,
- MorphNet veri seti için Özbekçe için hazır bir veri seti olmamasına rağmen başka diller için Universal Dependency Dataset içeriklerinin yazıldığı CoNNL-U formatında girişe olanak sağlanmıştır. Bu sayede daha farklı diller için farklı veri setleri ile MorphNet test edilebilmektedir. MorphNet başarı ölçümlemesi için başka dillerde ayrıca uygulama test edilmiştir,
- Özbekçe veri seti MorphNet yapısı gereği üç farklı parçaya rastgele bölünerek %80'i eğitim (train), %10'u geliştirme (development) ve %10'u test için kullanılmıştır. Kaynak kod üzerindeki düzenlemeler ile bu yapı değiştirilebilecek olsa da varsayılan olarak sunulduğu için değiştirilmemiştir. Referans çalışmadaki başarı ölçümlemesi bu doğrultuda olduğu için ayrıca bu nedenden dolayı da müdahale edilmemiştir,
- Belirli bir süre (4-8 saat) 8 GB GPU'ya sahip bir bilgisayar kullanılarak MorphNet modeli üzerinde eğitim (train) tamamlanarak Özbekçe metin, haber ve şiiir gibi içerikler test edilerek sonuçları paylaşılmıştır.

MorphNet sıralı yapay sinir ağı modeli ile Özbekçenin biçimbilimsel belirsizlik giderme için yapılan uygulamaların neticesinde daha fazla eğitim (train) yapılarak ve veri setinin boyutu artırıldıkça istenilen sonuçlara daha çok yaklaşılmıştır. Sonraki bölümde uygulama ile ilgili detaylar paylaşılmış olup, uygulama detayları ile elde edilen bulgular ve çıktılar paylaşılmıştır.

3.3. ÖZBEKÇENİN SÖZDİZİMSEL ÇÖZÜMLEMESİ

Tezin bu bölümünde Özbekçenin sözdizimsel çözümlemesi (syntactic parsing) üzerine yapılan araştırmalar, çalışmalar ve geliştirmeler sunulmuş olup, uygulama ve bulgular sonraki bölümlerde detaylandırılmıştır. Sözdizimsel çözümlemenin yapılabilmesi için Özbekçe dil grameri detaylı bir şekilde incelendikten sonra sözdizimsel çözümleme için kullanılan metot ve uygulamalar araştırılmıştır. Sözdizimsel çözümleme için kullanılan farklı yöntemler incelendikten sonra Özbekçe gibi sondan eklemeli dillerde ve özellikle Türki dillerde kullanılan sözdizimsel çözümleme yöntemleri üzerinde yoğunlaşılmıştır. Test edilen yöntem ve araçlardan en uygun olanı üzerinde geliştirmeler sağlanarak, Özbekçe grameri ve test verileri oluşturularak test ve uygulaması yapılmıştır. Burada yöntemin seçilmesinde, uygulanabilirlik, yöntemin sadeliği ve algoritmanın verimliliği göz önüne alınmıştır.

Bu çalışmada Özbekçenin sözdizimsel çözümlemesi için "Yukarıdan – Aşağıya (Top-Down)" sözdizimsel çözümleyici yaklaşımına sahip, morfolojik çözümlemeye (morphological analysis), bağlamdan bağımsız gramer (Context Free Grammar) kurallarına ve geçerlilik-doğrulama kurallarına (validation rules) dayanan bir çalışma referans (Zafer, 2011) alınarak bu çalışmadaki yöntem ve çözümleyici geliştirilerek kullanılmıştır. Bağlamdan bağımsız gramerin (CFG) Özbekçe ve karmaşık morfolojisi göz önüne alındığında en iyi çözüm olmadığı düşünülse de Türkçe ve Türkmence için yapılan referans çalışmanın sonuçları bu çalışmada tercih edilmesinde büyük rol oynamıştır. Ayrıca bağlamdan bağımsız gramerin basitliği, bilinirliği ve bu algoritmaları temel alan birçok verimli algoritmanın ve çözümleyicinin olması bu çalışmada tercih edilen yöntem olmasını sağlamıştır. Özbekçe için hazırlanan bağlamdan bağımsız gramer geliştirildikçe istenilen çıktılar alınmaya başlanmıştır. Referans çalışmada sunulan açık kaynak kodlu çözümleyici ve veri seti geliştirilmiş böylece bu çalışma işin beklenilen sonuçların daha da iyileştiği gözlemlenmiştir.

Türkçenin ve Özbekçe sözdizimini anlatan kaynaklar konu bütünlüğü, karşılaştırma ve uygulanacak yöntemin verimliliği için incelenmiştir. Şimşek (1987) kaynağı ve örnekleri referans alınarak Türkçe sözdizimi detaylandırılmıştır. Ayrıca referans çalışmada yer alan ve Türkçe dil kurallarının tanımlamasında kullanılan Karahan (2012) kaynağı da Türkçede sözdizimi, dil kuralları ve dil kurallarının tanımlanması, karşılaştırılması ve referans yöntemde uygulanması açısından incelenmiştir. Özbekçe ve Türkçe sözdizimi karşılaştırıldığında bazı küçük farklılıklar haricinde genel olarak aynı olduğu gözlemlenmiştir. Türkçe ve Özbekçe sözdizimini karşılaştıran Yaman (1994) ve Yaman (2000) kaynakları referans alınarak aşağıdaki benzerlik ve farklılık çıktılarına ulaşılmıştır.

- Özbekçe ve Türkçede cümle içerisindeki öğelerin sırası ve yeri aynıdır.
- Özbekçede ve Türkçede cümle öğelerinin sıralanışı Özne, Nesne ve Yüklem biçimindedir (Men kitobni o'qidim (Ben kitabı okudum)).
- Cümlede özne başta ve yüklem sonda yer almaktadır (Men o'qidim (Ben okudum)).
- Cümleler isim veya fiil cinsinden oluşan tek kelimeden oluşabilir (Keldim (Geldim)).
- Özbekçede cümleler söyleniş tarzına ve tonuna göre aynı cümle bildirme cümlesi, soru cümlesi ve ünlem cümlesi olabilir. Bu cümleler sonundaki noktalama işaretleri ile ayırt edilir. Örneğin; Kasal (Hasta), Kasal ? (Hasta mı?), Kasal! (Hasta!) cümleleri sırasıyla bildirme, soru ve ünlem cümleleridir.

- Aynı kelimelerden oluşan bazı cümleler durum ve ifadenin belirtilmesi için kullanılan noktalama işaretleri ile cümlenin farklı anlamlara sahip olmasına neden olabilir. Örneğin; Qancha kishi? (Kaç kişi?) ve Qancha kishi! (Nice/Çok kişi!) cümleleri aynı kelimelerden oluşmasına rağmen farklı anlamlara gelmektedir.
- Özbekçe ve Türkçede bazı edebi eserlerde (şiirler, destanlar, ağıtlar vb.) cümle öğelerinin yer ve sırası değiştirilebilir.

Türkçe ve Özbekçe sözdizimlerinde olmasa da gramer anlayışı ve metotlarda farklılık gözlenmektedir.

- Türkçede sıfat fiil tamlamasında iyelik ekleri tamlayana (sıfat fiil ekine) gelirken, Özbekçede tamlanana gelir. Örneğin; uyg'ongan paytlarim (uyandığım zamanlar).
- Özbekçede soru eki bitişik yazılırken, Türkçede ayrı yazılır. Örneğin; Sobir uning bolasimi? (Sobir onun çocuğu mu?).
- İstisnai bazı durumlarda (edebi eserlerde) belirtme hali eki ilgi hali görevini üstlenerek belirtili isim tamlaması oluşturabilmektedir.

Karşılaştırmalar sonrasında Türkçe ve Özbekçede söz diziminin birkaç küçük farklılık haricinde aynı olduğu söylenebilmektedir. Elde edilen sonuca göre Türkçe ve Türkmence için hazırlanan, Zafer (2011) çalışmasının ve çözümleme aracının geliştirilerek Özbekçe için uygulanmasının uygun olduğu kanaatine varılmıştır.

Özbekçe sözdizimsel çözümlemesi için Özbekçe gramer kuralları, çözümleyicinin desteklediği şekilde bağlamdan bağımsız gramer kuralları olarak hazırlanmış ve Özbekçe gramerine uygun geçerlilik – doğrulama kuralları hazırlanmıştır. Özbekçenin sözdizimsel çözümlemesinin yapılması için öncelikli olarak Özbekçenin gramerinin ve sözdiziminin detaylı bir şekilde incelenmesi gerekmektedir. Özbekçe söz dizimi ve grameri için bir çok yayın incelenerek dil kuralları bu doğrultuda hazırlanmıştır. Nurmanov ve diğ. (1992) ve Sulaymonov ve diğ. (1979) kaynakları ana referans alınarak Özbekçe dil kuralları hazırlanmıştır.

3.3.1. Bağlamdan Bağımsız Gramer

Bağlamdan bağımsız gramer (CFG), doğal dil işlemede dil tasarımı ve dilin kurallarını tanımlamak için kullanılan biçimsel ve matematiksel bir gramer türüdür. Bağlamdan bağımsız gramerin bir dil için tanımında temel gösterimi; A → b şeklinde olup, burada A devamlısının (nonterminal), b sonuncusuna (terminal) dönüşümü gösterilmektedir. Her bir kuralın aşağıdaki biçimde olması, bağlamdan bağımsız gramerin Chomsky normal biçiminde olduğu gösterir (Sipser, 1997) ve bu biçim doğal dillerin kelime ve cümle yapılarının tanımlanması için geliştirilmiştir.

$$A \rightarrow BC$$

$$A \rightarrow a$$

Yukarıda gösterildiği üzere a bir sonuncuyu (terminal) göstermektedir. A, B ve C devamlılarından (nonterminal), A başlangıç sembolünü göstermektedir. Herhangi bir bağlamdan bağımsız dil Chomsky normal biçiminde bağlamdan bağımsız gramer ile oluşturulabilmektedir.

Bağlamdan bağımsız gramer temel olarak dört özellikten meydana gelmelidir. Başlangıç sembolü (starting symbol), sonlular (terminals), devamlılar (nonterminals) ve bağlantılardan (relations) oluşur. İçerikten bağımsız bir gramer G ile gösterilerek tanım kümesi ve açıklamaları aşağıdaki gibidir.

$$G = (V, \sum, R, S)$$

- V devamlıları ifade eden sonlu bir kümedir,
- $\sum$, V'den ayrık sonlu bir kümedir ve sonluları ifade eder,
- R bağlantıları ifade eden sonlu bir kurallar dizisidir,
- $S \in V$'nin ve başlangıç sembolünü göstermektedir.

Bağlamdan bağımsız gramer sonlu kurallar dizisinden ve bağlantılardan oluşur. Buna göre sol tarafta devamlı (nonterminal) semboller, sağ tarafta sonlu (terminal) ya da devamlı (nonterminal) semboller bulunmaktadır. Aşağıdaki gibi tanımlanan bir dilde, şu gösterimler kullanılabilir ve çıktıları şu şekilde olabilir:

S → aSb

S → ab

G= ({S} , {a,b} , {S → aSb | ab} , S) gösterimi ile sunulan bu gramer ab, aabb, aaabbb, aaaabbbb, aaaaabbbbb gibi çıktılar (katarlar) üretebilir.

Özne, nesne ve yüklemden oluşan “Ahmed qora kitobni o’qidi (Ahmet siyah kitabı okudu)” cümlesi için temel bir bağlamdan bağımsız gramer örneği ve çözüm ağacı aşağıdaki gibidir.

S → VP

S → OBJ VP

S → SUB OBJ VP

SUB → Ahmed

SUB → qora

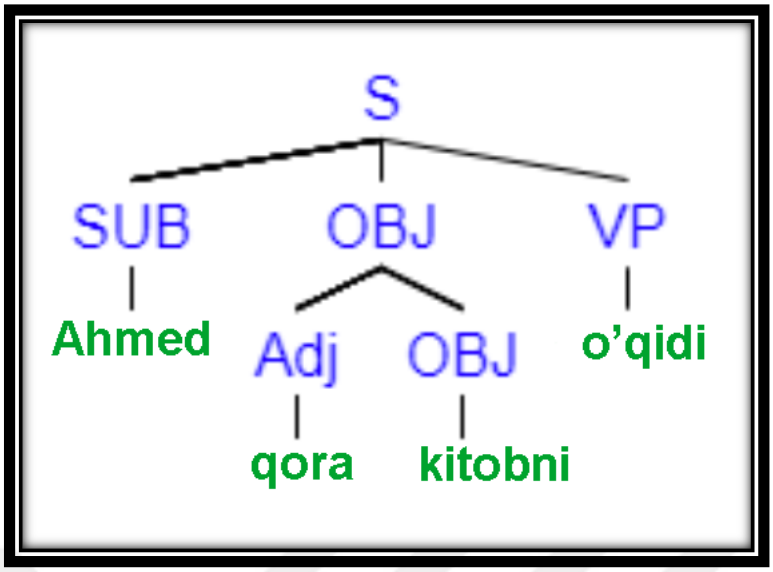
Adj → qora

OBJ → qora kitobni

OBJ → kitobni

VP → o’qidi

Yukarıda tanımlanan gramer ile üretilen cümlenin çözüm ağacı Şekil 3.4 ve çözüm ağacının parantez gösterimi aşağıdaki gibidir.



Şekil 3.4: Ahmed qora kitobni o'qidi (Ahmet siyah kitabı okudu) cümlesinin çözüm ağacı.

Çözüm ağacının parantez gösterimi aşağıdaki gibidir:

[S[SUB[Ahmed]][OBJ[Adj[qora]][OBJ[kitobni]]][VP[o'qidi]]]

Burada yukarıdan-aşağıya (top-down) bir çözümleme ile kök düğüm S'den başlayarak ağacı oluşturmaya ve yaprak düğümlerine kadar oluşturmaya devam etmektedir.

3.3.2. Özbekçe Sözdizimi ve Çözümlemesi

Bu bölümde Özbekçe sözdizimi detaylı olarak anlatılmış olup, sözdizimsel çözümlemeler örnekler ile sunulmuştur. Karahan (2012) çalışması referans alınarak kelime gruplarının sınıflandırılmasından, örneklendirilmesine bu bölüm bu kaynak doğrultusunda detaylandırılmıştır. Özbekçenin Türkçe ile benzerliği ve farklılıklarının bu çalışmaya yansıtılması için Yaman (2000) referans alınarak Özbekçe sözdizimi büyük bir hassasiyetle incelenmiştir.

Özbekçe kelime gruplarının sınıflandırılarak, her biri alt bölümlerde sözdizimsel olarak detaylandırılmış, Özbekçe kelime grupları bir önceki bölümde örnek olarak verilen cümledeki gibi CFG kuralları, ihtiyaç duyulan kadar örnek ve çözüm ağaçları ile verilmiştir. Çözüm ağaçları sonraki bölümlerde detaylı olarak anlatılan Nuve Parser uygulama aracı ile üretilmiş ve gösterilmiştir.

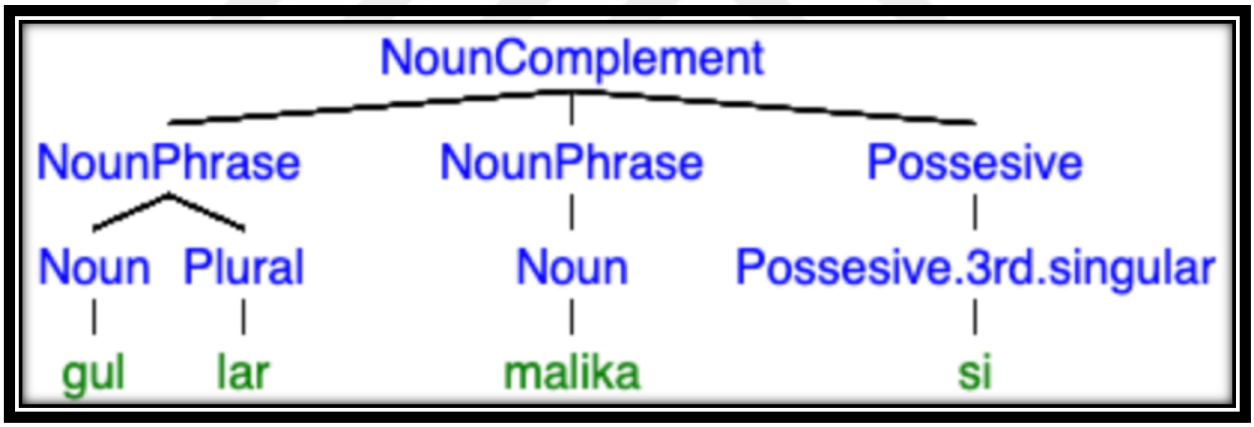
3.3.2.1. İsim Tamlaması

Özbekçede isim tamlaması (noun complement) belirtili ve belirtisiz isim tamlaması olarak iki veya daha fazla isim ya da isim soylu kelimelerin bir araya gelmesi ile oluşur. İsim tamlaması tamlayan ve tamlanan kelime gruplarının aldığı eke göre belirtili veya belirtisiz olarak adlandırılır. Tamlayan ilgi eki alması durumunda tamlama belirtili isim tamlaması, ek almadığı durumlarda belirtisiz isim tamlaması olarak adlandırılır. İstisnai bazı durumlarda (edebi eserlerde) belirtme hâli eki ilgi hali fonksiyonunu üstlenerek belirtili isim tamlaması oluşturabilmektedir. İsim tamlamasının bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

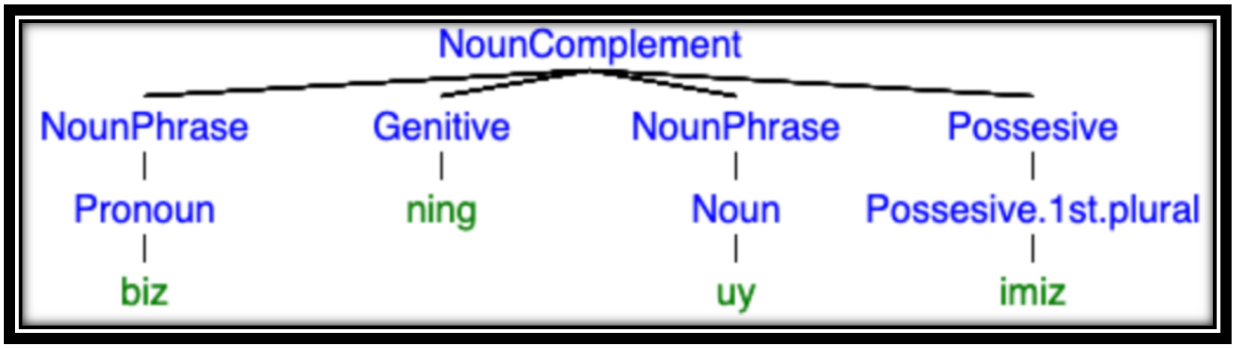
NC → NP NP Poss

NC → NP gen NP Poss

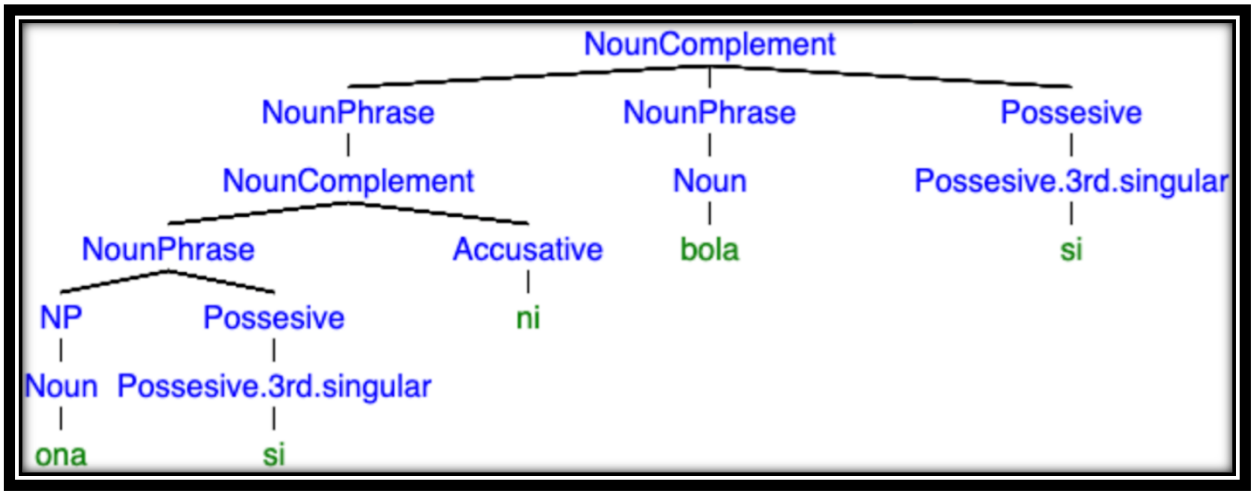
Poss → p1s | p2s | 3s | p1p | p2p | p3p



Şekil 3.5: Gullar malikasi (Çiçeklerin melikesi) – Belirtisiz isim tamlaması.



Şekil 3.6: Bizning uyimiz (Bizim evimiz) – Belirtili isim tamlaması.



Şekil 3.7: Onasini bolasi (Annesinin çocuğu) – Belirtili isim tamlaması (istisna).

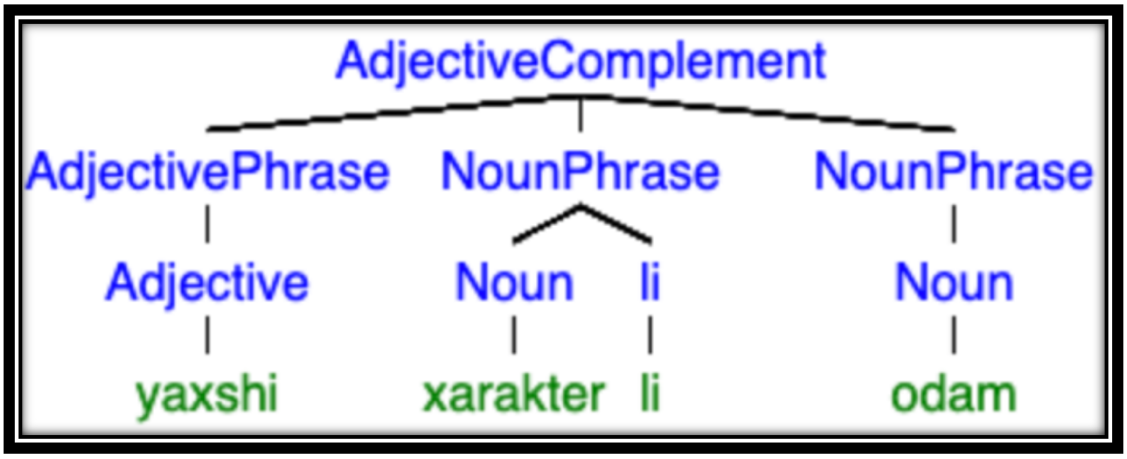
3.3.2.2. Sıfat Tamlaması

Tamlayanı sıfat ya da sıfat grubu tamlananı isim ya da isim grubu olan tamlama çeşididir. Sıfat tamlamasının (Adjective complement) bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

JC → JP NP

JC → JP cm JC

JC → JC cm NP



Şekil 3.8: Yaxshi xarakterli odam (İyi karakterli insan) – Sıfat tamlaması.

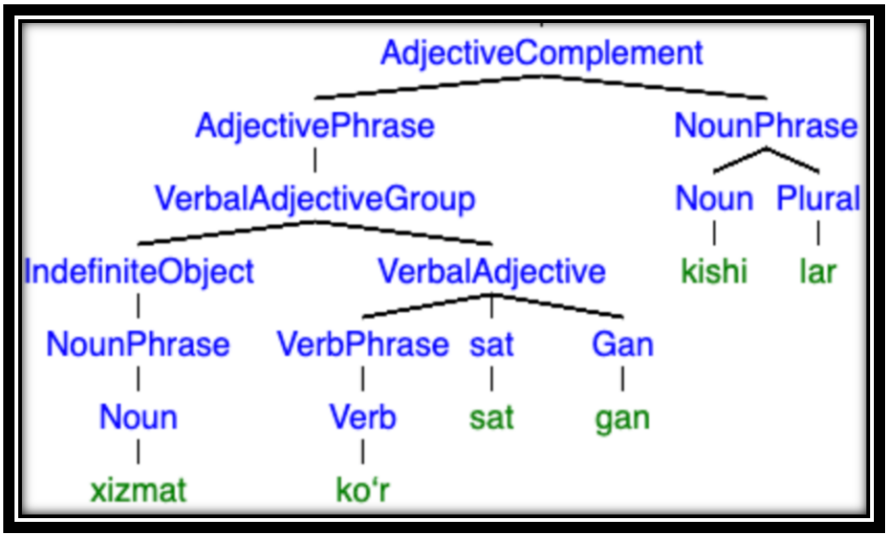
3.3.2.3. Sıfat-Fiil Grubu

Özbekçede sıfat-fiil grubu (verbal adjective group), sıfat-fiil ve belirtili ya da belirtisiz bir tamlayandan oluşmaktadır. Sıfat-fiil grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

VJG → VJ

VJG → IOBJ VJ

VJG → DOBJ VJ



Şekil 3.9: Xizmat ko’rsatgan kishilar (Emek sarfeden kişiler) – Sıfat-fiil grubu.

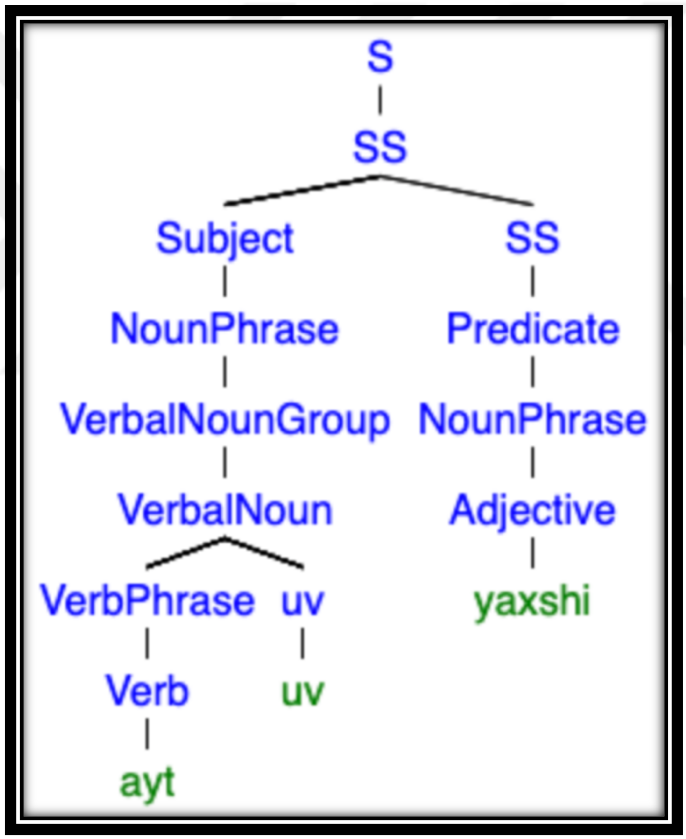
3.3.2.4. İsim-Fiil Grubu

İsim-fiil grubu (verbal noun group), bir isim fiilden ve belirtili ya da belirtisiz bir tamlayandan oluşmaktadır. İsim-fiil grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

VNG → VN

VNG → IOBJ VN

VNG → DOBJ VN



Şekil 3.10: Aytuv yaxshi (Söylemek iyidir) – İsim-fiil grubu.

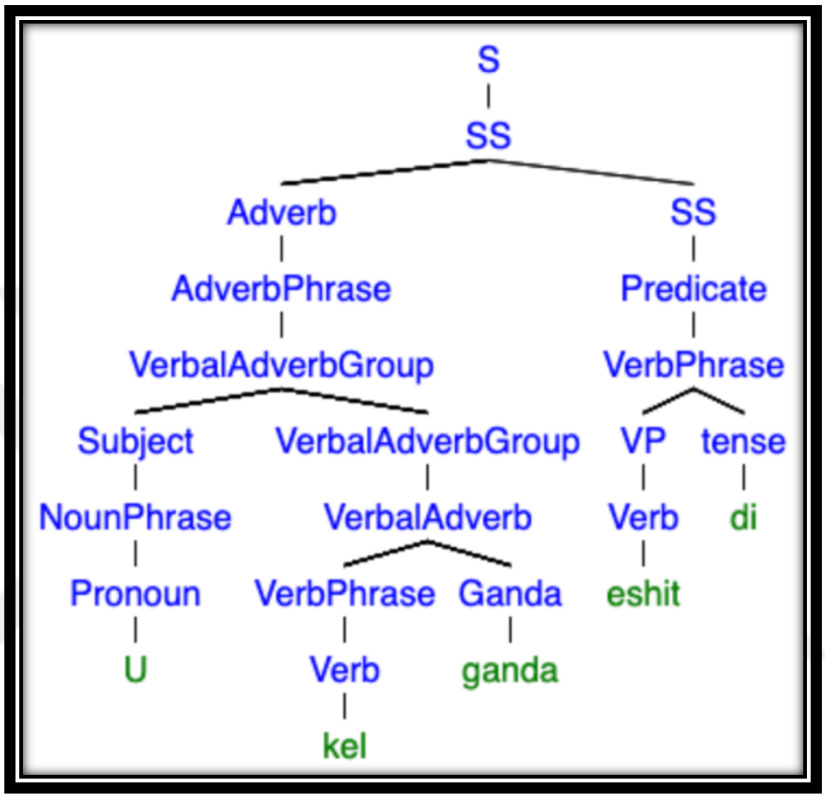
3.3.2.5. Zarf-Fiil Grubu

Zarf-fiil grubu (verbal adverb group), bir zarf fiilden ve belirtili ya da belirtisiz bir tamlayandan oluşmaktadır. Zarf-fiil grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

VAG → VA

VAG → IOBJ VA

VAG → DOBJ VA

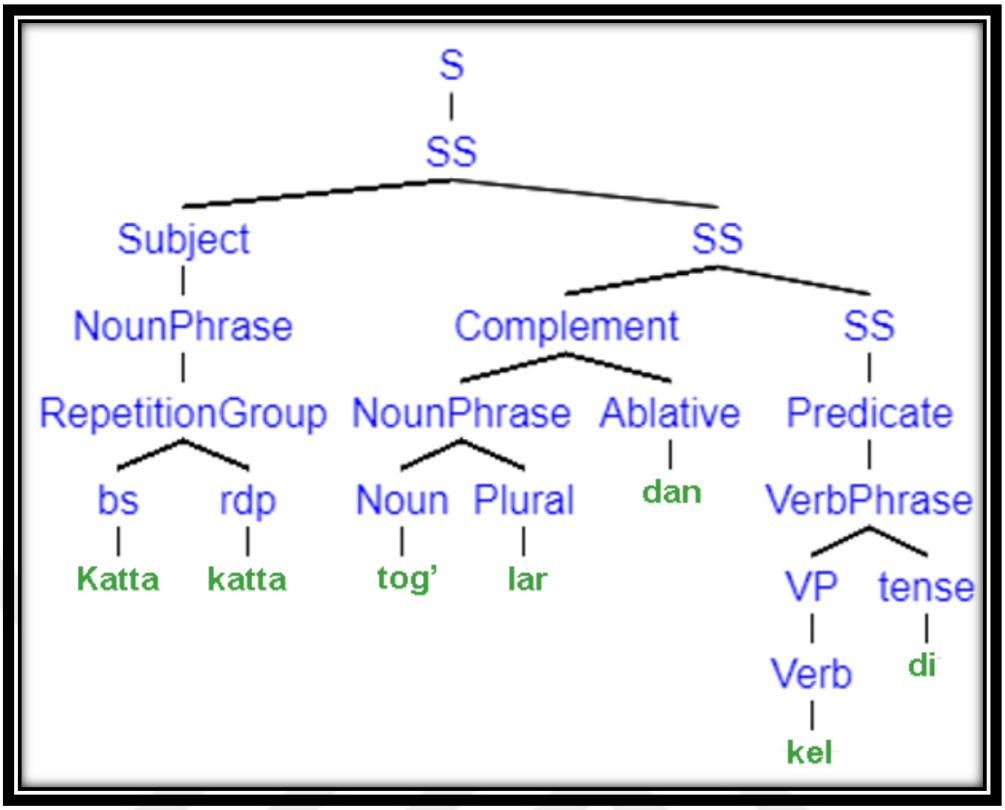


Şekil 3.11: U kelganda eshitdi (O geldiğinde işitti) - Zarf-fiil grubu.

3.3.2.6. Tekrar Grubu

Pekiştirme, tasvir etme ve detaylandırma için sıklıkla kullanılan ve eş görevli kelimelerin oluşturduğu gruptur. Tekrar grubu (reduplication group) bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

RG → base redup



Şekil 3.12: Katta-katta tog'lardan keldi (Büyük büyük dağlardan geldi).

3.3.2.7. Edat Grubu

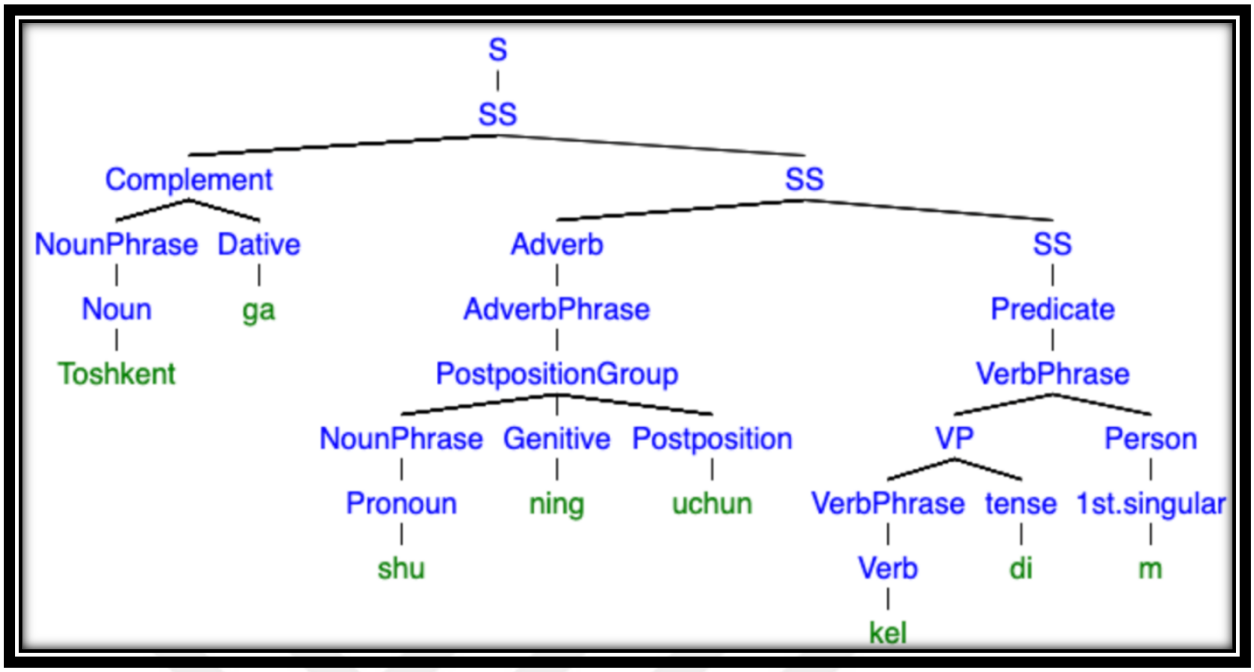
Özbekçede edat grubu (postposition group), bir isim öbeği ile bir edat/çekim edatından oluşmaktadır. Edat grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

PG → NP posp

PG → NP dat posp

PG → NP gen posp

PG → NP abl posp



Şekil 3.13: Toshkentga shuning uçkun keldim (Taşkent'e şunun için geldim) – Edat grubu.

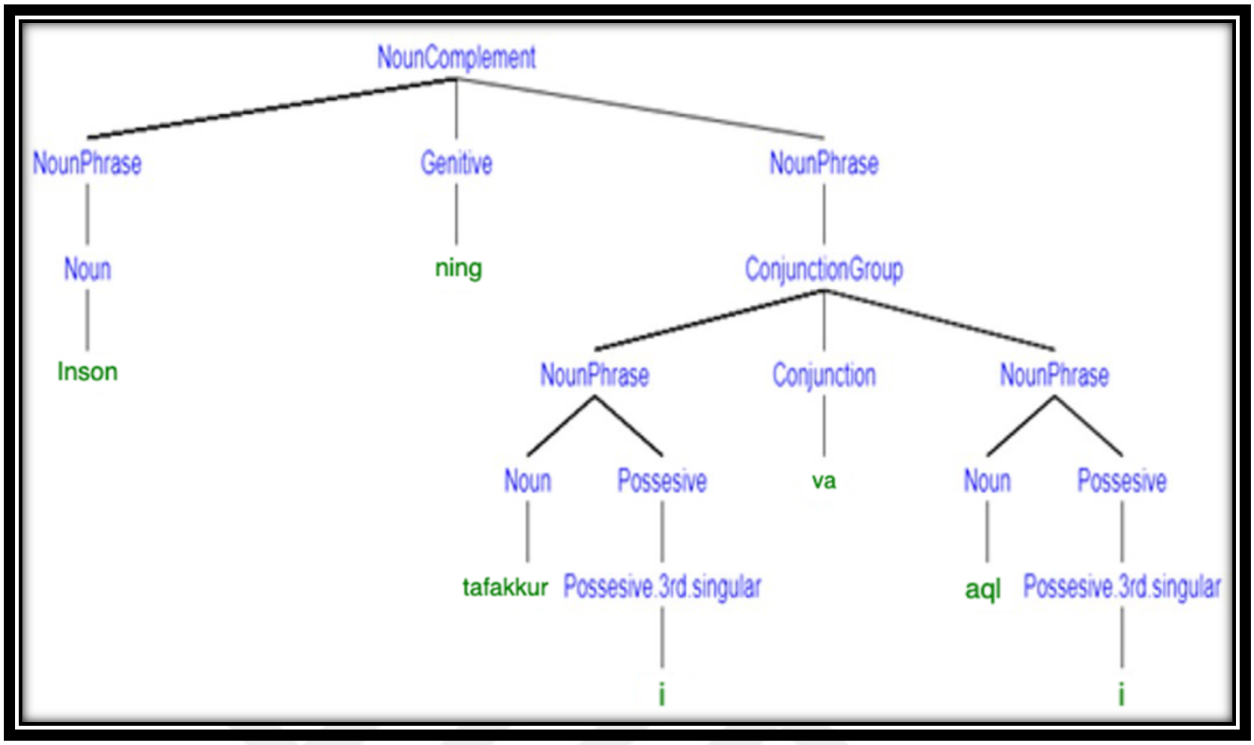
3.3.2.8. Bağlama Grubu

Bağlama grubu (conjuction group), iki ya da daha fazla isim öbeğinin bağlaçlar ile bağlanması sonucu oluşur. Bağlama grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

CG → NP conj NP

CG → NP cm CG

CG → NP ins posp



Şekil 3.14: Insonning tafakkuri va aqli (İnsanın tefekkürü ve aklı) – Bağlama grubu.

3.3.2.9. Unvan Grubu

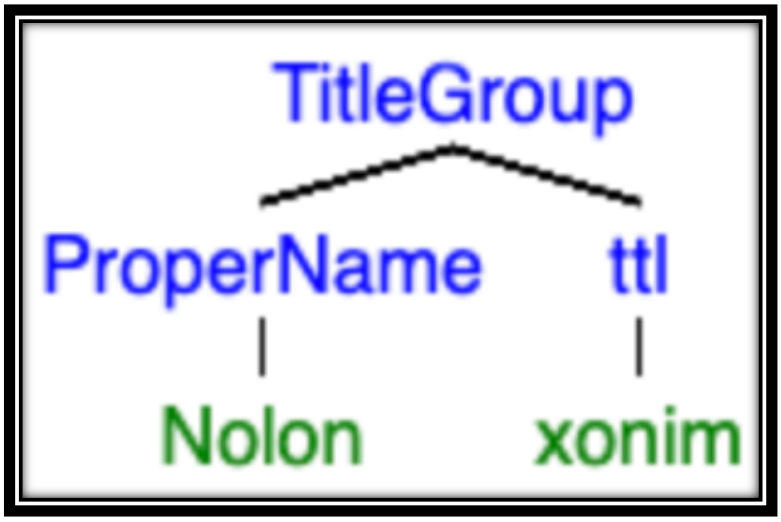
Özbekçede unvan grubu (title group), bir şahıs ve bir unvandan oluşan kelime grubudur. Unvan grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

TG → prop ttl

TG → ttl prop

TG → PNG ttl

TG → ttl PNG



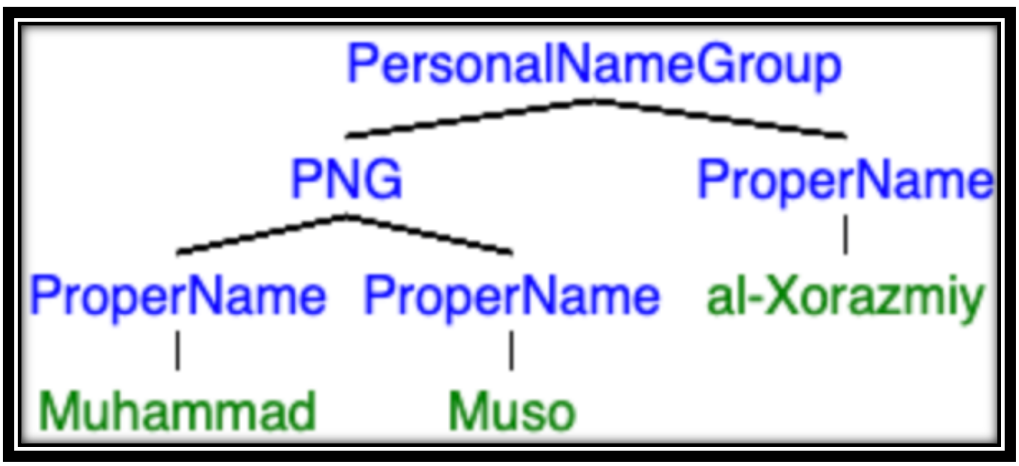
Şekil 3.15: Nolon xonim (Nalan hanım) – Unvan grubu.

3.3.2.10. Özel İsim Grubu

Özel isim grubu (person name group), iki ya da daha fazla özel isimden oluşan kelime grubudur. Özel isim grubu bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

PNG → prop prop

PNG → PNG prop



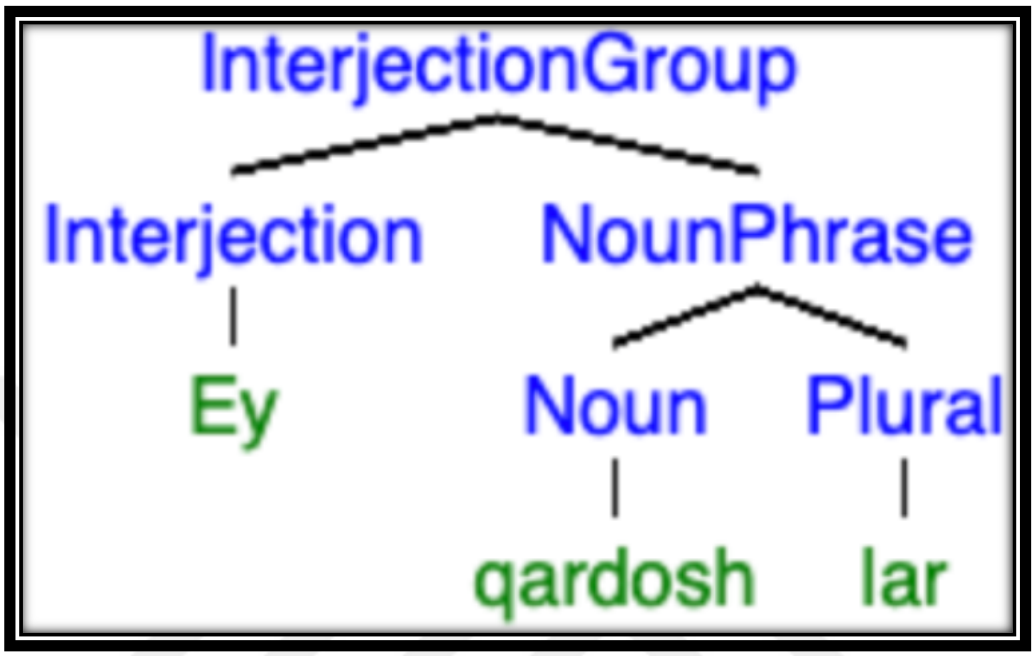
Şekil 3.16: Muhammad Muso al-Xorazmiy (Muhammed Musâ el-Harezmi).

3.3.2.11. Ünlem Grubu

Özbekçede ünlem grubu (interjection group), bir ünlem ve bir veya birden fazla isim öbeğinden oluşan kelime grubudur.

IG → intj NP

IG → IG cm NP



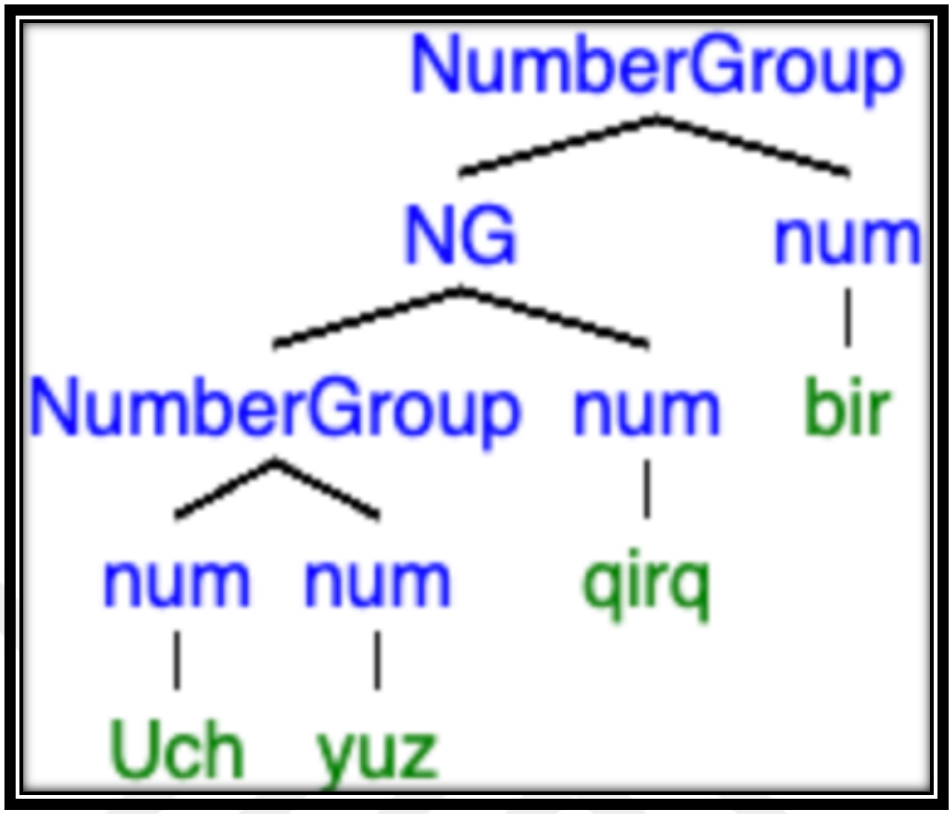
Şekil 3.3.17: Ey gardoshlar (Hey kardeşler)! – Ünlem grubu.

3.3.2.12. Sayı Grubu

Sayıların bir araya gelerek matematiksel olarak birler, onlar, yüzler vb. basamak sistemi ile oluşturduğu gruptur. Sayı grubu (number group) cümle içerisinde isim veya sıfat olabilir. Sayı grubunun bağlamdan bağımsız gramer ve çözüm ağacı gösterimleri aşağıdaki gibidir.

NG → num num

NG → NG num



Şekil 3.18: Uch yuz qirq bir (Üç yüz kırk bir) – Sayı grubu.

3.3.2.12. Birleşik Fiil Grubu

Özbekçede birleşik fiil grubu (compound verb group), iki ya da daha fazla kelimeden (fiil ya da fiil grubu) oluşan, bir eylemi çeşitli bakımlardan tasvir eden kelime grubudur.

CVG → NP XVP

XVP → auxv

XVP → XVP psv

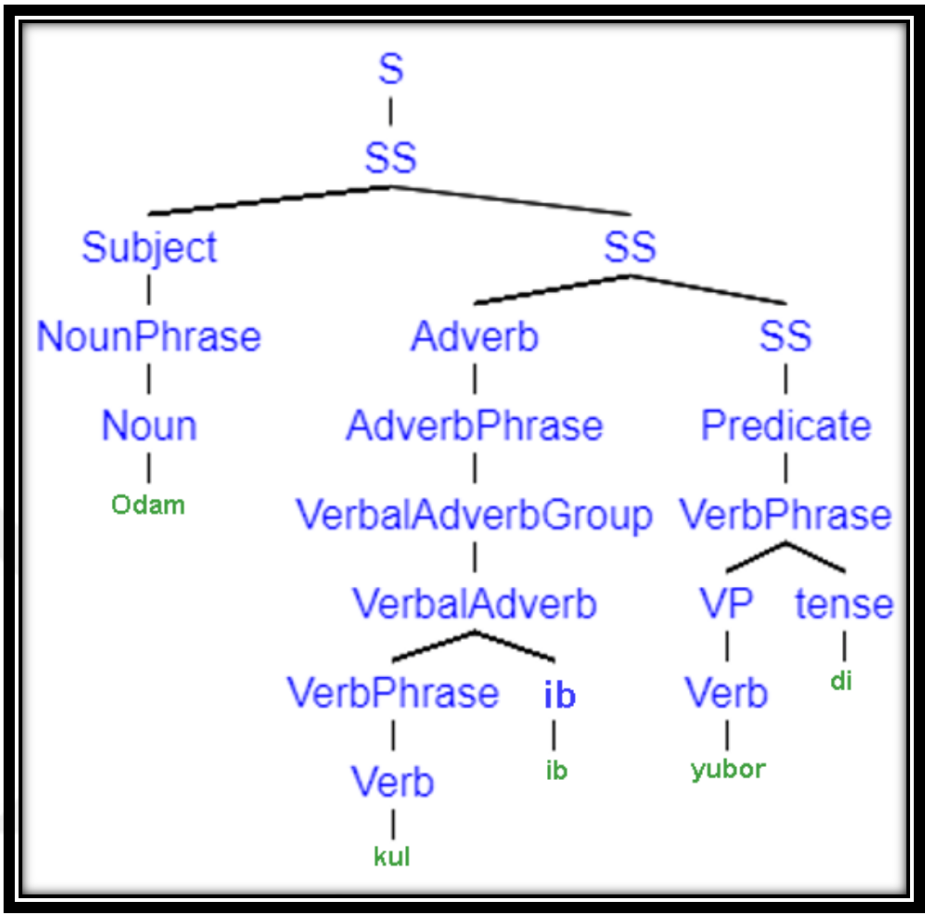
XVP → XVP neg

XVP → XVP tense

XVP → XVP Prs

XVP → XVP aux

XVP → XVP mood



Şekil 3.19: Odam kulib yubordi (Adam gülüverdi) – Birleşik fiil grubu.

4. BULGULAR

Tezin bu bölümünde, Özbekçenin biçimbilimsel ve sözdizimsel çözümlemesinin uygulamalarına ve bu uygulamalar sonrası elde edilen bulgulara değinilmiş, uygulama aşamaları ve çıktıları detaylandırılmıştır.

Özbekçenin biçimbilimsel çözümlemesi için bu çalışma için geliştirilen ve detaylandırılan Nuve Studio[2] kullanılmış olup, Özbekçenin biçimbilimsel belirsizlik gidermesi için MorphNet modeli ve uygulama çevresi kullanılarak Özbekçe için geliştirmeler sağlanmıştır. Özbekçenin sözdizimsel çözümlemesi için Nuve Parser kullanılmış olup, bu çalışma ve uygulama ortamındaki ihtiyaçlar için yazılım geliştirilerek istenilen çıktılar alınmış ve sonuçları paylaşılmıştır.

Özbekçenin biçimbilimsel çözümlemesi ve uygulaması ile biçimbilimsel belirsizlik giderme konuları ve uygulamaları ilgili bölümde detayları ile sunulmuştur. Sonraki bölümde Özbekçenin sözdizimsel çözümlemesinin detayları ve uygulama aşamaları ile çıktıları paylaşılmıştır.

4.1. NUVE STUDIO İLE ÖZBEKÇENİN BİÇİMBİLİMSEL ÇÖZÜMLEMESİ

Bu tezde Özbekçenin iki düzeyli morfolojisi detayları ile sunulmuş olup bir önceki bölümde kullanılan yöntem detaylandırılmıştır. Bu bölümde ilgili yöntemin uygulamasının yapılması için gerekli çalışmalar, uygulama örnekleri ve uygulama esnasında karşılaşılan özel durumlar ile istisnalar sunulmuştur.

Bir dilin biçimbilimsel çözümlemesinin yapılabilmesi için temel olarak dilin yazım (ortografik) kurallarına, biçim dizgesi(morfotaktik) kurallarına, dil sözlüğüne (kökler ve ekler sözlüğü gibi) ve dilin tanımlanarak çözümlenebilmesi için bir morfolojik çözümleyiciye ihtiyaç duyulmaktadır.

Özbekçenin morfolojisi Nuve Studio[2] biçim bilimsel çözümleyicisi kullanılarak test edilmiş ve uygulaması yapılmıştır. Nuve Studio(Küçük, 2021) bu tez kapsamında geliştirilen, Nuve (Zafer, 2018) çatısı ve kütüphanelerinin geliştirilerek Özbekçe ve birçok Türki dil için

[2] www.nuvestudio.com

kullanılmasını sağlayan bir uygulama aracıdır. Nuve Studio geliştirilirken açık kaynak kodlu, biçim bilimsel çözümleme ve makine çevirisi için kullanılan Nuve çatısı ve kütüphaneleri geliştirilerek nesne tabalı PHP diline çevirilmiş, Özbekçe için ihtiyaç duyulan geliştirmeler yapılarak online www.nuvestudio.com web adresinden erişime açılmıştır. Bu geliştirmeler yapılırken uygulama aracının online erişime açılması gerekliliği hissedilmiş, Türki diller üzere çalışan araştırmaların uygulama çevresini kullanmasına olanak tanınmıştır. Uygulama geliştirmelerinin uygulamayı geliştirmesi için kaynak kodları paylaşılmıştır (Küçük, 2021).

Herhangi bir dilin Nuve Studio ile biçim bilimsel çözümlemesinin yapılabilmesi için dilin kök sözlüğü (root lexicon), ek sözlüğü (suffix lexicon), yazım kuralları (orthography) ve biçim dizgesinin (morphotactics) Nuve Studio tarafından kabul edilen dosya biçime uygun olarak hazırlanması gerekmektedir. Özbekçe için öncelikli olarak birden fazla sözlük taranarak kök kelime sözlüğü oluşturulmuştur. Özbekçe gramer kuralları ve ek yapısı öğrenilerek ek sözlüğü, yazım kuralları dosyası ve biçim dizgesi kurallarını içeren biçim dizgesi (morfotaktik) dosyası oluşturulmuştur. Örnek olarak, Özbekçe için hazırlanan uygulama dosyalarına http://www.nuvestudio.com/downloads/uz_UZ/uz_UZ.zip linkinden erişilebilir olup, bu dosyaların hazırlanma aşaması ve detayları sonraki alt bölümlerde paylaşılmıştır.

4.1.1. Nuve Studio ile Uygulama için Kökler Sözlüğünün Hazırlanması

Özbekçenin biçim bilimsel çözümlemesinin Nuve Studio ile yapılabilmesi için öncelikli olarak uygulama dosyalarının Özbekçe için hazırlanması gerekmektedir. Bunun için birçok Özbekçe sözlük incelenerek bu verilerin kök sözlüğüne aktarılması için dijital olarak taratılan kaynakların aktarımı için bir çevirim programı hazırlanarak kök kelime (root lexicon) sözlüğü hazırlanmıştır.

Kök kelime sözlüğü (root lexicon) comma-seperated variables (CSV) dosya formatında(roots.csv) hazırlanarak bu dosya içerisinde root, surfaces, lex, active, id, flags ve rules sütunları Nuve Studio formatına uygun olarak yer almaktadır. Bu alanların dışında bazı özel durumlar, test ve geliştiricinin okunabilirliğini artırmak gibi amaçlar ile ek sütunların açılması ile esneklik sağlanmıştır.

Özbekçe sözlüğünde yer alan kelimeler farklı sözlükler kullanılarak uygulamanın kök kelime sözlüğü geliştirilmiş, uygulama için 30.000 kök kelime Nuve Studio biçimine uygun olarak tanımlanmıştır. Buna göre her bir kelime root sütununda, kelimenin metin içerisindeki farklı

tüm hallerinin tanımları surfaces sutününda, kelimenin türü (POS) id sutününda, özel durumların tanımlanması flag sütununda ve ortografik kuralları rules sütununda tanımlanmaktadır. Tanımlanan bu sütunlar kelimenin özelliklerini (attribute) göstermenin yanında Nuve Studio programının doğru çalışması için ilişkileri içermektedir. Ortography dosyasında tanımlanan kuralların ekler ve kelimeler ile eşleştirilmesinin yanında kök kelime sözlüğünde yer alan rules sütunu özel durumların tanımlanması için kullanılmaktadır. Nuve Studio için hazırlanan bir kelimenin morphotactics dosyası ve root dosyası arasındaki ilişkisi id sütunu ile sağlanmaktadır. Buna göre yazım kurallarına uygun olarak kök ve ek sözlüğü, yazım ve biçim dizgesi kuralları için morphotactics ve ortography dosyalarının oluşturulması gerekmektedir. Nuve Studio için hazırlanan kökler (roots) sözlüğünden bir kesit Tablo 4.1'de sunulmuştur.

Tablo 4.1: Nuve Studio için hazırlanan kökler sözlüğünden bir kesit.

root	**surfaces**	**lex**	**active**	**id**	**flags**	**rules**
avzo	avzoy			ISIM	isim	TUREME_Y
axtar				FIIL	fiil	
ayir	ayr			FIIL	fiil	DUSME_SON_UNLU

4.1.2. Nuve Studio ile Uygulama için Ekler Sözlüğünün Hazırlanması

Özbekçenin biçim bilimsel çözümlemesinin Nuve Studio ile yapılabilmesi için hazırlanması gereken bir diğer dosya Özbekçe eklerini içeren ekler sözlüğüdür. Bunun için Özbekçe gramer kaynaklarından faydalanarak Özbekçe ek yapısı çıkarılmıştır. Özbekçede yer alan ekler ile ekler sözlüğü (suffix lexicon) hazırlanmıştır.

Ekler sözlüğü (suffix lexicon) de comma-seperated variables (CSV) dosya formatında (suffixes.csv) hazırlanarak bu dosya içerisinde id, lexical form, type, group, rules ve surfaces sütunları Nuve Studio formatına uygun olarak yer almaktadır. Bu sütunların (attribute) haricinde başka sütunlarda yer almakta olup, Nuve Studio uygulamasının fonksiyonelliğini artırmak ve özel durumların çözümlenmesi gibi amaçların yanında ilerideki geliştirmelere esneklik sağlaması amacıyla bu şekilde tasarlanmıştır.

Nuve Studio formatına uygun şekilde 370 ek, ekler sözlüğüne eklenmiştir. Eklenen her bir ek sırasıyla, ek id sütununda tanımlanan ek tanımı ile orthography ve morphotactics dosyaları ile ilişkilendirilmektedir. Id ekin adını tanımlamanın yanında biçim dizgesi ve yazım kuralları ile ilişkilendirmeyi sağlayan bir değere sahiptir. Sonraki sütunda lexical form yer almakta olup, bu sütun ekin yapısal biçimde gösterimini ifade etmektedir. Type sütunu ekin işlevsel türünü göstermekte, group sütunu ekin kategori grubunu göstermektedir. Rules sütununda ekin hangi kurallarda yer aldığını göstermektedir. Surfaces sütunu ekin yapısal olarak gösterildiği lexical form sütundaki değerinin yüzeyde aldığı tüm değerleri göstermektedir. Bu değerlerin ve sütunların haricinde eklenen sütunlar test, özel durum veya gelecekteki çalışmaların esneklik sağlaması düşünülerek için oluşturulmuştur. Örneğin isInflectional sütunu bu eklerin çekim ya da yapım eki olma özelliğine göre raporlama ve test gibi ihtiyaçlar için ayırıcı bir değer oluşturması gerekliliği ile hazırlanmış olup, biçim bilimsel çözümleme için değer bulundurması zorunlu bir sütun değildir. Nuve Studio için hazırlanan ekler (suffix) sözlüğünden bir kesit Tablo 4.2'de sunulmuştur.

Tablo 4.2: Nuve Studio için hazırlanan ekler sözlüğünden bir kesit.

id	**lexical form**	**type**	**group**	**rules**	**surfaces**
IC_COGUL_LAR	lar	IC	COGUL	dek	lar
IC_IYELIK_BEN_HM	im	IC	IYELIK	DUSME_UNLU	im, m, yim
IC_YONELME_GA	Ga	IC	HAL	DONUSUM_G, TUREME_Y	nga, ga, g'a, ka, qa

4.1.3. Nuve Studio ile Uygulama için Yazım Kurallarının Hazırlanması

Bir dilin yazım (ortografik) kuralları, o dilde kelimelere gelen eklerin, kelimeyi türetmesini, çekimlemesini ve kelime ya da ekte meydana gelen değişimlerin kurallar çerçevesinde gerçekleştirilmesini sağlayan kuralların bütünüdür. Bu kurallara uygun olarak kelimenin gövdesinde ya da ekinde değişimler meydana gelmektedir.

Kelime gövdesine gelen eklerin eklenme sırası, düzeni, şekli ve kelimede ya da ekte oluşan değişimler, kelimenin biçim bilimsel çözümleme aşamasında da aynı kurallar çerçevesinde

çözümleneceği için bir dilin biçim bilimsel çözümlemesi ve kelime köküne eklerin getirilmesi ile türetilmesi yazım kurallarına uygun olarak yapılmaktadır. Biçim dizgesi sonraki alt bölümde anlatılacak olup, belirli sıra ve düzene uygun bir şekilde gelen eklerin ve kökün değişimi yazım kuralları ve biçim dizgesi kurallarına bağlıdır.

Nuve Studio ile hem biçim bilimsel çözümleme hem de kelime köküne eklenecek ek türlerinin belirtilmesi ile yeni kelimenin üretilmesi yapılabilmektedir. Bu üretim ve çözümleme aşamasında bu işlemi ilgili dil için tanımlanan yazım (ortografi) kuralları çerçevesinde, Nuve Studio'da yer alan ortography dosyasında tanımlanan kurallara uygun olarak yapmaktadır.

Özbekçenin yazım kuralları iki düzeyli kurallar olarak hazırlanmış ve detayları önceki bölümde paylaşılmıştır. Bu kurallar Özbekçenin grameri detaylı olarak araştırılıp öğrenildikten sonra hazırlanmıştır. Nuve Studio içerisinde yer alan ortography dosyasına bu kurallar uygun formatta tanımlanmıştır. Ortography dosyası bir extensible markup language (XML) dosyası (orthography.xml) olup iki düzeyli kuralları içermektedir.

Her bir iki düzeyli kural bu dosya içerisinde rule etiketleri arasında tanımlanır. Her bir rule etiketi id ve phase niteliklerini taşır. Id (identify) niteliği kök (roots.csv) ve ek(suffixes.csv) sözlükleri ile ilişkilendirilerek ilgili kök veya ek ile ilgili yazım kurallarının tanımlanmasını sağlamaktadır. Phase (faz) niteliği ilgili kuralın uygulanma sırasını tanımlamaktadır. Phase niteliğinin sahip olduğu değere göre (değerin küçük olması öncelikli olduğu anlamına gelir) sıralanarak, öncelikli kuralların daha önce kelime veya eke uygulanıp sonra diğer kuralların uygulanmasını sağlamaktadır. Description (tanımlama) etiketi ilgili kuralın tanımını ve açıklamasını içermekte olup, daha okunabilir bir yapının kurgulanması ve çalışmayı hazırlayan kişinin kural açıklamalarını hatırlamasına yardımcı olmak için oluşturulmuştur. Conditions (durumlar) etiketleri kural içerisinde eğer(if) durumlarını oluşturmaktadır. Kural içerisinde eğer şu durum gerçekleşirse şu şekilde sonuçlandır ifadeleri conditions etiketlerinin içerisinde yer alan her bir condition etiketinde tanımlanmaktadır. Transformation (dönüşüm) etiketi kelimede ya da ekte bir harfin veya bir ses grubunun içerisinde yer alan operand niteliklerindeki değerler ile değiştirilmesini sağlamaktadır. Birden fazla durum (conditions) birleştirilerek sonucun elde edilebilmesi için And ve Or mantıksal operatörleri kullanılarak ve/veya işlemleri yapılmaktadır. Bu operatör değerleri flag niteliğine atanarak ilgili işlemler seçilmektedir.

Condition etiketi morpheme, operatör ve operand niteliklerini içermektedir. Morpheme niteliği kelime içerisinde yer alan morfeme erişerek işlemin yapılacağı morfemin seçilmesini sağlamaktadır. İlgili morfemin kelime içerisindeki konumu ya da morfemlerin birbirleri ile olan sırasına ve ilişkisine göre ilgili morfeme erişim sağlanmaktadır. Operator ve operand nitelikleri ilgili morfeme erişmek için durum kontrollerinin sağlanması ve erişim sonrası işlemlerin yapılması için farklı değerler almaktadır. Transformation (dönüşüm) etiketi morpheme, action, operands ve flag niteliklerinden oluşmaktadır. Action niteliği ekleme (insert), silme (delete) ve değiştirme (replace) gibi değerler alarak kelimenin ya da ekin yüzeysel formda değiştirilmesini dönüştürülmesini sağlayan aksiyonların tanımlanmasını sağlamaktadır. Diğer nitelikler condition etiketinde tanımlanan özellikler ile aynı fonksiyonelliklere sahiptir.

Özbekçenin iki düzeyli yazım (ortografik) kuralları Nuve Studio formatına uygun bir şekilde hazırlanarak ortography.xml dosyasına eklenmiştir. Düzenli kuralların yanı sıra, birçok özel durum ve istisna kurallar ayrıca tanımlanarak kelime kökünde veya ekte tanımlanan kurallara göre çözümlemenin ve üretmenin gerçekleşmesi sağlanmıştır. Nuve Studio için hazırlanan yazım kurallarından (orthography.xml) bir kesit Şekil 4.1'de sunulmuştur.

```
<rule id="DONUSUM_D" phase="2">
  <description>
     Cs sert sessizlerinden biri ile biten gövdeye D={d, t} ile başlayan bir ek geldiğinde,
     D meta harfi yüzeyde t sessizi ile gösterilir. Cs sert sessizlerinden biri ile
     bitmeyen gövdelere eklenen ekin başındaki D meta harfi d sessizi ile gösterilir.
  </description>
  <transformation morpheme="This" action="Replace" operandOne="D" operandTwo="t" flag="all">
    <conditions flag="Or">
      <condition morpheme="Previous" operator="LastLetterEquals" operand="prfstkqchshlnsh" />
    </conditions>
  </transformation>
  <transformation morpheme="This" action="Replace" operandOne="D" operandTwo="d" flag="all" />
</rule>
```

Şekil 4.1: Nuve Studio için hazırlanan yazım kurallarından bir kesit.

4.1.4. Nuve Studio ile Uygulama için Biçim Dizgesi Kurallarının Hazırlanması

Özbekçenin isim(nominal) ve fiil(verbal) köklü biçim dizgesi bir önceki bölümde detayları ile paylaşılmış olup, bu detaylar uygulama aşamasında kullanılarak Nuve Studio için biçim dizgesi kuralları hazırlanmıştır. Biçim dizgesi, morfemlerin sıralanmasına ilişkin düzen kısıtlamalarını temsil eder. Morfemlerin birbirini nasıl takip edeceklerini, sıralamalarını ve düzenini tanımlayan kurallar dizisidir. Bu kuralların Nuve Studio içerisinde tanımlanması biçim bilimsel

çözümlemede sıra ve düzenin oluşturulmasını sağlamaktadır. Nuve Studio programında bu kuralların tanımlanması için morphotactics dosyası yer almaktadır.

Morphotactics dosyası da bir extensible markup language (XML) dosyası olup ekleri ve eklerin birbirleri ile olan ilişkilerinin sıralandığı bir yapıya sahiptir. Özbekçenin biçim dizgesi sonlu durum makineleri ile ifade edildiği gibi bir sonraki gelecek ekin durumu, Nuve Studio'da XML ile tanımlanarak gruplandırılmaktadır.

Suffix Group etiketiyle tanımlanan ekler aynı grupta yer alıp ortak özelliklere sahiptir. Burada bir gruplandırma ile ortak özelliklerin ortak fonksiyonelliğe sahip olması sağlanmıştır, böylece her ek grubu ortak olarak tanımlanarak, her ekin ayrı ayrı daha uzun tanımlanmasının önüne geçilmiştir. Suffix Group etiketiyle ortak özelliklere sahip olan her bir ek suffix etiketi ile bu grubun içine dahil edilir. Örneğin iyelik ekleri bir grup altında toplanmaktadır. Source etiketi başlangıç durumunu(state) sunmakta olup, id niteliğine sahiptir. Id niteliği kaynağın(source) türünü(isim, fiil gibi) göstermektedir. Source etiketinin içerisinden sonraki durumları(states) göstermek için target ve target group etiketleri yer almaktadır. Target group yukarıda anlatılan suffix group ile aynı mantık doğrultusunda oluşturulmuş olup ortak özelliklerin daha kısa yazılması için oluşturulmuştur. Source etiketi içerisinde yer alan target ve target group etiketiyle geçişi tanımlanan sonraki durumlar, sırasıyla yazılarak gövdeye gelme sıraları belirlenmiştir. Örneğin, bir isim başlangıç durumu olabilir ve bu durumda source etiketinde id niteliği isim olarak tanımlanır. Sonraki durum (ara durum) çoğul eki almış bir isim olabilir. Özbekçede çoğul eki bir tane olduğu için çoğul eki sıradaki duruma geçiş için target etiketi ile tanımlanabilir. İsim köklü biçim dizgesi paradigması gereği iyelik ekleri çoğul ekinden sonra gelmektedir. Buna göre çoğul ekinden sonra tanımlanacak durumlara geçişte iyelik ekleri kullanılabilir. Birden fazla iyelik eki olduğu için target group etiketiyle tanımlaması yapılmaktadır. Target ve target group etiketleri son duruma (final state) ulaşıncaya kadar eklenip tanımlamasının yapılması gerekir. Eksik yapılan bir tanımlama sonucun doğruluğuna etki edeceği için Özbekçe grameri detaylı olarak çalışılarak, Nuve Studio ile uygulama için biçim dizgesi kuralları morphotactics dosyasında dikkatle hazırlanmıştır.

Özbekçenin biçim dizgesi(morfotaktik) kuralları, Nuve Studio formatına uygun olarak sunulmuştur. Özel ve istisna durumlar, istenilen sonuçlara ulaşmak için ayrıca tanımlanmıştır. Nuve Studio için hazırlanan biçim dizgesi kurallarından (morphotactics.xml) bir kesit Şekil 4.2'de sunulmuştur.

```
<?xml version="1.0" encoding="utf-8"?>

 <morphology lang="Uz-UZ">
    <suffixGroup name="IYELIK">
          <suffix>IC_IYELIK_BEN_Hm</suffix>
          <suffix>IC_IYELIK_BIZ_Hmiz</suffix>
          <suffix>IC_IYELIK_SEN_Hng</suffix>
          <suffix>IC_IYELIK_SIZ_Hngiz</suffix>
          <suffix>IC_IYELIK_SIZ_Hnglär</suffix>
          <suffix>IC_IYELIK_SIZ_läring</suffix>
          <suffix>IC_IYELIK_O_(s)H</suffix>
          <suffix>IC_IYELIK_ONLAR_läri</suffix>
    </suffixGroup>
    …
    <graph>
          <source id="ISIM">
                <target id="IC_COGUL_lär"></target>
                <targetGroup id="IYELIK"></targetGroup>
                 <targetGroup id="HAL"></targetGroup>
                 …
          </source>
          …
     </graph>
 </morphology>
```

Şekil 4.2: Nuve Studio için hazırlanan biçim dizgesi kurallarından bir kesit.

4.1.5. Nuve Studio ile Özbekçe Biçimbilimsel Çözümleme Örnekleri

Tezin bu bölümünde Nuve Studio[3] uygulama ortamında test edilen Özbekçe örnekler ve biçim bilimsel çözümlenmiş çıktıları sunulmaktadır. Bir kelimenin birden fazla çözümünün olması biçimbilimsel belirsizlik giderme bölümde ve uygulama aracında çözümlenecek olup, örnekleri ilgili bölümde yer almaktadır.

Kökler sözlüğünde tanımlanan kelimelerin ve ekler sözlüğünde ilişkilendiren eklerin, yazım kuralları ve biçim dizgesi kuralları çerçevesinde Nuve Studio tarafından verilen çözümleri bazı kısaltmalar ile aşağıda verilmiştir. Örneklerdeki kelimeler Nuve Studio text alanına yazılarak çıktıları ile örneklerin anlaşılması için Türkçe anlamları ile paylaşılmıştır.

[3] www.nuvestudio.com

Olmangiz

olma+ngiz	[İSİM(olma)+2.ÇOĞUL ŞAHIS İYELİK(ngiz)]	(elmalarınız)
ol+ma+ngiz	[FİİL(ol)+F.İ.YAPIM EKİ(ma)+ 2.ÇOĞUL ŞAHIS İYELİK(ngiz)]	(bir şey alma)
ol+ma+ngiz	[FİİL(ol)+OLUMSUZLUK(ma)+2.ÇOĞUL ŞAHIS EMİR(ngiz)]	(alma)

Bo’ldimi

bo'l+di+mi	[FİİL(bo’l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS+SORU(mi)]	(oldu mu?)
bo'l+di+mi	[FİİL(bo’l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS+SORU(mi)]	(böldü mü?)

Bekmi

bek+mi	[İSİM(bek)+3.TEKİL ŞAHIS+SORU(mi)]	(bey mi?)
bek+mi	[SIFAT(bek)+3.TEKİL ŞAHIS+SORU(mi)]	(kapalı mı?)

Sirdagi

sir+da+gi	[İSİM(sir)+BULUNMA(da)+İLGİ(gi)]	(sırdaki)
sir+da+gi	[İSİM(sir)+BULUNMA(da)+İLGİ(gi)]	(boyadaki)
sir+da+gi	[İSİM(sir)+BULUNMA(da)+İLGİ(gi)]	(peynirdeki)

Keltirilgan

kel+tir+il+gan	[FİİL(kel)+ETTİRGEN(tir)+EDİLGEN(il)+GEÇMİŞ ZAMAN(gan) +3.TEKİL ŞAHIS]	(getirildi)
kel+tir+il+gan	[FİİL(kel)+ETTİRGEN(tir)+EDİLGEN(il)+ SIFAT-FİİL(gan)]	(getirilmiş)

Qotma		
qot+ma	[FİİL(qot)+F.İ.YAPIM EKİ(ma)]	(sıska)
qot+ma	[FİİL(qot)+OLUMSUZLUK(ma)]	(sert olma)
qot+ma	[FİİL(qot)+F.F.YAPIM EKİ(ma)]	(sertleşme)
qotma	[İSİM(qotma)]	(sıska)

Nuve Studio ile biçimbilimsel çözümlemesi esnasında ortaya çıkan özel durumlar ve istisnalar bir sonraki alt bölümde örnekler ile detaylandırılarak, tanımlanan kurallara göre beklenen çıktıların istisna durumları paylaşılmıştır.

4.1.6. Özel Durumlar, İstisnalar ve Önerilen Çözümler ile Özbekçenin Nuve Studio ile Biçimbilimsel Çözümlemesi

Bu bölümde, Özbekçenin biçimbilimsel çözümlemesinin Nuve Studio ile test ve uygulaması yapılırken karşılaşılan özel durumlar ve istisnaların yanında bu durumlar için önerilen çözümler sunulmuştur. Biçimbilimsel belirsizlik giderme bölümünde bir kelimenin birden fazla biçimbilimsel çözümü üzerinde durularak kelimenin cümle içerisindeki kullanımına göre istenilen biçimbilimsel çözümleme tek bir çözüm olarak elde edilmektedir.

Bazı özel durumlar ve istisnaların çözümlenebilmesi için tanımlanan grup kurallardan ayrılması ve ayrıca istisna olan durumun tanımlanması gerekmektedir. İstenilen çözümün elde edilebilmesi için birçok özel durum ve istisna için özel kurallar tanımlanmıştır. Özel durumlar ve istisnalar örnekler ile sunulurken, örnek kelimelerin biçim bilimsel çözümlemesi ve ek yapısı ayrıca gösterilmiştir. Özel durumlar, kelimenin cümle içerisindeki kullanımı ve özel durumu göstermek için cümleler ile sunulmuştur. Kelimeler, beklenen ve istisna durumun gerçekleştiği ya da gerçekleşmediği kelimenin doğru gösterimi ile örneklendirilmiştir. Özel durumlar ve istisnalar aşağıda örneklerle detaylandırılmıştır.

1. Bazı sayılara -dan veya -lab ekleri eklenince gövde zarfa dönüşür. Bu sayılara eklenen aynı ek bazı durumlarda gövdenin zarfa dönüşmeyerek sayı olarak kalmasına neden

olur. Çekim eki istisnai durumlarda yapım eki gibi davranır, -dan çekim eki yapım eki gibi davranmıştır. Aşağıdaki örneklerin ilkinde -dan ekini alan sayı gövdesi zarfa dönüşürken, ikinci örnekte sayı olarak kalmıştır.

Biçimbirimler	**Yapısal**	**Kelime**	**Cümle**
bir+dan	SAYI(bir)+ÇIKMA(dan) (zarf – aniden)	birdan	Birdan kelip goldi. (Birden gelip kaldı.)
bir+dan	SAYI(bir)+ÇIKMA(dan) (sayı – bir)	birdan	Birdan birni kamaytsa nol goladi. (Birden bir çıkarsa sıfır kalır.)

Bu özel durumun çözümlenmesi biçimbilimsel belirsizlik giderme bölümünde sağlanarak cümlede içerisindeki kullanımına göre doğru çözüm belirlenmektedir.

2. İşaret zamirlerine eklenen bazı hal ekleri(yönelme, bulunma ve çıkma) ve bazı çekim eklerinin arasına n sessizi ve ni harfleri türer. Üçüncü çoğul şahıs ekleri işaret zamirlerine eklenmesi durumunda istisnai olarak ni türemesi gözlenmez.

Biçimbirimler	**Yapısal**	**Beklenen**	**Doğru**
shu+da	ZAMİR(shu)+BULUNMA(da)	shuda	shunda(şunda)
bu+ga	ZAMİR(bu)+YÖNELME(ga)	buga	bunga(buna)
shu+si	ZAMİR(shu)+3.TEKİL ŞAHIS İYELİK EKİ(si)	shusi	shunisi(şunusu)
shu+lar+i	ZAMİR(shu)+ÇOĞUL(lar)+ 3.ÇOĞUL ŞAHIS İYELİK EKİ(i)	shunilari	shulari(şunları)

İşaret zamirlerine eklenecek üçüncü çoğul şahıs ekleri için ayrıca istisna bir kural tanımlanmıştır. Yukarıda belirtilen işaret zamirlerine ve üçüncü çoğul şahıs eklerine bu kuralın ilişkilendirmesi ile Nuve Studio ile çözümleme sağlanmıştır.

3. Özbekçe'de geçmiş zaman eki sadece -di ekidir. Geçmiş zaman eki ünsüz uyumundan dolayı herhangi bir ses değişimine uğramaz.

Biçimbirimler	Yapısal	Beklenen	Doğru
ket+di	FİİL(ket)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	ketti	ketdi(gitti)
top+di	FİİL(top)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	topti	topdi(bulundu)

Özbekçede C_s sert sessizlerinden biri ile biten gövdeye D={d, t} ile başlayan bir ek geldiğinde D meta harfi yüzeyde t sessizi ile gösterilir. Ancak bu durum istisna olarak geçmiş zaman eki için geçerli değildir. Nuve Studioda bu kontrol gerçekleştirilerek eğer ek geçmiş zaman eki ise değişim gerçekleştirilmeden çözüm yapılmaktadır.

4. Özbekçede bazı kelimelerin yazılışı aynı fakat anlamları farklıdır. Farklı eklerin bu gövdelere eklenmesi sonrası yazılışları aynı olsa bile farklı özel durumlara ve anlamlara sahip olabilir. Ayrıca, aynı ekler eklense bile bu gövdelerin anlamları ve morfolojik yapısı farklı olabilir. Bu kelimelerin hangisinin ne anlama geldiği ancak cümle veya konuşma dilinde anlaşılabilir olup, morfolojik belirsizliğe yol açmaktadır. Bu kelimelerin cümle içinde doğru morfolojik çözümlenmesi için morfolojik belirsizlik giderme yöntemlerinin kullanılması gerekmektedir. Uygulama esnasında farklı kurallar tanımlanarak, bu kelimeler için özel durumlar oluşturulmuştur.

Biçimbirimler	Yapısal	Sonuç
bo'l+di	FİİL(bo'l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	bo'ldi(böldü)
bo'l+di	FİİL(bo'l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	bo'ldi(oldu)
bek+mi	İSİM(bek)+3.TEKİL ŞAHIS+SORU(mi)	bekmi?(Bey mi?)
bek+mi	SIFAT(bek)+3.TEKİL ŞAHIS+SORU(mi)	bekmi?(Kapalı mı?

qot+ma	FİİL(qot)+F.İ.YAPIM EKİ(ma)	qotma(sıska)
qot+ma	FİİL(qot)+OLUMSUZLUK(ma)	qotma(sert olma)
qot+ma	FİİL(qot)+F.F.YAPIM EKİ(ma)	qotma(setleşme)
qotma	İSİM(qotma)	qotma(sıska)

5. Bir gövdenin son harfi g, g', k ve q sessizlerinden biriyse ve -Gan geçmiş zaman veya -Ga hal eklerinin biri gövdeye eklenirse, ekin ilk sessizi ve gövdenin son sessizi k veya q sessizine dönüşür. Gövdenin son harfleri g, g', k ve q sessizlerinden biri değilse bu durum gerçekleşmez. Özbekçede normalde, eğer bir gövde g, g', k ve q sessizlerinden biriyle biter ve G sessizi ile başlayan bir ek alırsa, sadece G sessizi k ve q sessizine dönüşür.

Biçimbirimler	**Yapısal**	**Beklenen**	**Doğru**
teg+gan	FİİL(teg)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	teggan	tekkan(değmiş)
bo'g'+gan	FİLL(bo'g')+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	bo'g'gan	boqqan(besledi)
tik+gan	FİLL(tik)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	tikgan	tikkan(dikilmiş)
uq+gan	FİLL(uq)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	uqgan	uqqan(uyanmış)
ket+gan	FİLL(ket)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	ketgan	ketgan(gitmiş)
bosh+ga	İSİM(bosh)+YÖNELME(ga)	boshga	boshga(başka)

Bazı kelimeler için oluşan istisna durumlar kök kelime sözlüğünde o kelime için tanımlanan kural ile çözümlenmiştir.

6. Bir gövdenin son harfi y, r, sh ve n sessizlerinden biriyse ve -Dir ettirgen fiil eki eklenirse, ekin ilk sessizi t(tir) olarak görünür. Eğer gövde herhangi bir başka sessizle biterse, bu durum gerçekleşmez. Gövdenin son harfinin l sessizi olması durumunda,

-Dir eki -tir olarak görünür. Sadece bir istisna kelimede bu durum görünmez. Kel(gel) kelimesi ve istisnai durumu aşağıda sunulmuştur.

Biçimbirimler	**Yapısal**	**Beklenen**	**Doğru**
aylan+Dir	FİİL(aylan)+ETTİRGEN(dir)	aylandir	aylantir(dönüştür)
chiqar+Dir	FİİL(chiqar)+ETTİRGEN(dir)	chiqardir	chiqartir(çıkarttır)
ko'pey+Dir	FİİL(ko'pey)+ETTİRGEN(dir)	ko'peydir	ko'peytir(çoktur)
kel+Dir	FİİL(kel)+ETTİRGEN(dir)	keldir	keltir(getir)
ol+Dir	FİİL(ol)+ETTİRGEN(dir)	oldir	oldir(öldür)

Bazı kelimeler için oluşan istisna durumlar kök kelime sözlüğünde o kelime için tanımlanan kural ile çözümlenmiştir.

7. Bu istisna, -chAQ son ekiyle ilgilidir. Duruma göre ek -choq veya -chak olarak gözlemlenir. Bir kelimede x,q ve g' sessizleri olduğunda, kelimedeki arka sıralı ünlüler kesinlikle arka sıralı ünlüler gibi davranır. Diğer durumlarda, arka sıralı ünlüler kelimedeki ön sıralı ünlüler olarak sunulur. Bu durumlar aşağıda örneklendirilmiştir.

Biçimbirimler(Morphemes)	**Yapısal(Lexical)**	**Sonuç**
qo'zic+chAQ	İSİM(qo'zic)+İ.İ.YAPIM EKİ(chAQ)	qo'zichoq (kuzucuk)
o'rgim+chAQ	İSİM(o'rgim)+İ.İ.YAPIM EKİ(chAQ)	o'rgimchak(örümcek)

Bu gibi istisna durumlarda, yazım kurallarına ayrıca bir kural eklenerek ve bu kuralın kök kelime sözlüğünde eşleştirmesi yapılarak çözüm sağlanmıştır.

4.2. MORPHNET İLE ÖZBEKÇE İÇİN BİÇİMBİLİMSEL BELİRSİZLİK GİDERME

Tezin bir önceki bölümünde detaylandırılan Özbekçenin biçimbilimsel çözümlemesinin (morphological disambiguation) uygulama aşamaları ve çıktıları bu bölümde detaylandırılmıştır. Dayanık ve diğ. (2018) çalışmasında detayları paylaşılan ve bu çalışma için referans alınan çalışmanın, Akyürek (2019) tarafından açık kaynak kodlu olarak sunulan uygulama aracı düzenlenerek, geliştirilerek ve Özbekçe için uyarlanarak biçimbilimsel belirsizlik giderme için kullanılmıştır.

Özbekçenin biçimbilimsel belirsizlik gidermesi için 30.000 kelimeden oluşan bir veri seti kullanılarak, bu veri setinin 24.000'i eğitim (train), 3.000'i geliştirme(dev) ve 3.000'i test için kullanılarak sonuçları paylaşılmıştır. Çalışmanın ölçümlenmesi için ayrıca 300 kelime harici tutulmuştur.

MorphNet modeli girdi olarak aldığı metnin içerisindeki kelimelerin doğru biçimbilimsel çözümlemesini çıktı olarak üreten, bir sıralı (seq2seq) tekrarlayan yapay sinir ağı (RNN) modelidir. Yapay zeka uygulamalarının derlenmesi için ekran kartı GPU'su kullanılmakta olup programın çalıştırılması ve eğitimin (train) yapılması için Nvdia GPU gereklidir.

Özbekçenin biçimbilimsel belirsizlik giderme uygulaması için detayları alt bölümlerde yer alan aşağıdaki işlemler ve hazırlıklar yapılmıştır.

- MorphNet veri setinin hazırlanması için Nuve Studio ile biçimbilimsel çözümlemesi yapılacak veri setinin için kaynakların belirlenmesi,
- Kaynaklar belirlendikten sonra kaynaklarda içeren kelimelerin Nuve Studio kök kelime sözlüğüne eklenmesi ve bu kelimelerden hali hazırda sözlükte olan kelimelerin elenmesi,
- Nuve Studio ile kaynakta yer alan metinlerin biçim bilimsel çözümlemesinin yapılması,
- Biçimbilimsel çözümlemeleri yapılan kelimelerin MorphNet veri seti girdi formatına uygun hale getirilmesi,
- Birden fazla çözümün yer aldığı veri setinde en doğru çözümün elle işaretlenmesi ve ilk sıraya yerleştirilmesi,
- Veri seti bölünerek yeterli eğitimin (train) yapılması,

- Eğitim sonrası test yapılarak sonucun, en doğru biçimbilimsel çözümün çıktısının alınması ve değerlendirme,

Yukarıdaki işlemler ile Özbekçenin biçimbilimsel belirsizlik giderme işlemi tamamlanmış olup, bu bölümün alt bölümlerinde uygulama aşamaları, uygulama öncesi yapılan hazırlıklar, uygulama örnekleri ve çıktıları detayları ile verilmiştir.

4.2.1. MorphNet ile Uygulama için Veri Setinin Hazırlanması

Özbekçenin biçimbilimsel çözümlemesi sonrası ortaya çıkan belirsiz durumların giderilmesi için biçimbilimsel belirsizlik içeren kelimelerin cümle ve metinleri ile MorphNet veri setine eklenerek biçimbilimsel belirsizlik giderme sonrası tek bir çıktı beklenmiştir. Uygulamanın anlamlı veriler üretebilmesi için yeterli miktarda veri ve eğitimin (train) sağlanması gerekliliğinden dolayı Özbekçe hikaye, makale, haber ve şiir gibi edebi eserler ayrıca biçimbilimsel olarak çözümlenerek veri setine eklenmek için hazırlık yapılmıştır. MorphNet uygulama başarısının artması için yeterli eğitim ve veri önem arz etmektedir. Veri seti boyutu ve eğitim (train) süresi artırıldıkça çıktıların başarı oranı ayrıca artmıştır.

Nuve Studio ile biçimbilimsel çözümlemesi tamamlanan ve biçimbilimsel belirsizlik giderme için MorphNet uygulamasına verilen kelimeler haricinde veri setinin artırılması için bir çok yazılı metin incelenmiştir. Bu metinlerin öncelikli olarak biçimbilimsel çözümlemesinin yapılması için, metinde yer alan kelimelerden Nuve Studio kök kelime sözlüğünde olmayan kelimeler kök kelime sözlüğüne eklenmiştir. İlk aşamada 13.000 ve sonraki aşamada 30.000 kelimelik bir veri seti oluşturulmuştur. Bu veri setinin boyutunun artırılma ihtiyacı MorphNet başarı oranını artırma ihtiyacından ortaya çıkmıştır. Bu doğrultuda Nuve Studio kök kelime sözlüğünde tanımlanan kelimelerin biçimbilimsel çözümlemesi gerçekleştirilmiştir.

Nuve Studio çıktı ve MorphNet girdi formatları aynı olmadığı için ara bir çevirim programı (script) yazma ihtiyacı ortaya çıkmıştır. Örnek olarak Shirin nega uxlamayapsan (Shirin neden uyumuyorsun)? soru cümlesini veri setine eklemek istediğimizde öncelikli olarak bu cümleyi Nuve Studio ile biçimbilimsel çözümlemesini gerçekleştirerek aşağıdaki çıktılar alınmaktadır.

Shirin

Shirin/ISIM_OZEL

nega

ne/ZAMIR_SORU IC_HAL_YONELME_Ga

nega/ISIM

ne/ISIM IY_ISIM_AQ IY_ISIM_ä

ne/ISIM IY_ISIM_AQ IY_FIIL_ä

ne/ISIM IY_ISIM_Ga

ne/ISIM IY_FIIL_Ga

uxlamayapsan

uxla/FIIL OLUMSUZLUK_mä FC_ZAMAN_SIMDIKI_yäp EKFIIL_SAHIS_SEN_sän

uxla/FIIL FY_ISIM_Im IY_FIIL_ä FC_ZAMAN_SIMDIKI_yäp EKFIIL_SAHIS_SEN_sän

uxla/FIIL FY_FIIL_mä FC_ZAMAN_SIMDIKI_yäp EKFIIL_SAHIS_SEN_sän

Cümlesi çözümlendiğinde bazı kelimelerin birden fazla biçimbilimsel çözümünün bulunduğu gözlemlenmiştir. Ancak Nuve Studio çıktı formatı MorphNet girdi formatı için uygun değildir. MorphNet veri seti girdi formatı olarak Universal Dependency Dataset içeriklerinin yazıldığı CoNNL-U formatını ya da varsayılan olarak tanımlanmış TRDataset formatını kullanmaktadır. Bunun için Özbekçe veri seti girdi formatı TRDataset formatına çevrilmiştir. Universal Dependency Dataset içerikleri kontrol edildiğinde CoNNL-U formatında Türkçe dahil birçok dil için veri seti yer alsa da Özbekçe için veri seti bu tezin hazırlandığı tarihte yer almamaktadır. Özbekçe için bir veri seti yer almış olsaydı daha fazla veri ile test edebilmek için bu veri seti ayrıca kullanılacaktı. Bu nedenle varsayılan veri seti formatı TRDataset tercih edilmiştir. Buna göre yukarıda Nuve Studio çıktısı verilen cümlenin MorphNet girdi formatı aşağıdaki gibidir.

<S>

Shirin Shirin+ISIM_OZEL

nega ne+ZAMIR_SORU+IC_HAL_YONELME_Ga nega+ISIM
ne+ISIM+IY_ISIM_AQ+IY_ISIM_ä ne+ISIM+IY_ISIM_AQ+IY_FIIL_ä
ne+ISIM+IY_ISIM_Ga ne+ISIM+IY_FIIL_GA

uxlamayapsan

uxla+FIIL+OLUMSUZLUK_mä+FC_ZAMAN_SIMDIKI_yäp+EKFIIL_SAHIS_SEN_sän
uxla+FIIL+FY_ISIM_Im+IY_FIIL_ä+FC_ZAMAN_SIMDIKI_yäp

+EKFIIL_SAHIS_SEN_sän

uxla+FIIL+FY_FIIL_mä+FC_ZAMAN_SIMDIKI_yäp+EKFIIL_SAHIS_SEN_sän

</S>

Cümle halinde eklenen kelimeler sentence(cümle) anlamına gelen taglar(etiketler) arasına yazılmaktadır. Her bir kelimenin varsa birden fazla biçimbilimsel çözümü boşluk bırakılarak yazılmaktadır. Kelimeden sonra sunulan ilk çözüm bu cümle için ilgili kelimenin en doğru biçimbilimsel çözümlemesidir. MorphNet veri seti girdi formatı olarak aşağıdaki Şekil 4.3'te gösterilen CoNNL-U formatında da veri girişi kabul etmektedir.

```
# sent_id = 7700.train
# text = Oyunun afişte kalması için başarıyla oynanması gerekir .
1       Oyunun  oyun    NOUN    _       Case=Gen|Number=Sing|Person=3    3       nsubj   _       _
2       afişte  afiş    NOUN    _       Case=Loc|Number=Sing|Person=3    3       obl     _       _
3       kalması kal     NOUN    _       Case=Nom|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3    7       obl     _       _
4       için    için    ADP     _       _       3       case    _       _
5       başarıyla       başarı  NOUN    _       Case=Ins|Number=Sing|Person=3    6       obl     _       _
6       oynanması       oyna    NOUN    _       Case=Nom|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3    7       csubj   _       _
7       gerekir gerek   VERB    _       Aspect=Hab|Mood=Gen|Number=Sing|Person=3|Polarity=Pos|Tense=Pres|VerbForm=Fin       0       root    _       _
8       .       .       PUNCT   _       _       7       punct   _       _
```

Şekil 4.3: Oyunun afişte kalması için başarıyla oynanması gerekir, cümlesinin CoNNL-u formatında gösterimi.

Nuve Studio biçimbilimsel çıktı formatını MorphNet girdi formatına çeviren ara bir program(script) ile otomatik çevirim sağlandıktan sonra veri setinin elle işaretleme aşamasına geçilmiştir. Buna göre veri setinde yer alan ve eğitimde kullanılması planlanan kelimelerin içinde bulundukları cümle içerisindeki en doğru biçimbilimsel çözümleri işaretlenerek ilk çözüme taşınmıştır. Böylece Özbekçe için TRDataset formatına uygun veri seti hazırlanmıştır. Şekil 4.4'te bulunduğu cümlede "Ben evimi döşemeden, mobilya almadan evleneceğim, hatun

alacağım” anlamına gelen “men mebelsiz xotin olaman (ben mobilyasız hatun alacağım)” cümlesini içeren MorphNet veri setinden bir kesit sunulmaktadır.

```
<S>
men     men+ZAMIR_SAHIS men+ISIM
mebelsiz        mebel+ISIM+IY_SIFAT_siz mebel+ISIM+EKFIIL_SAHIS_SIZ_siz
xotin   xotin+ISIM
olaman  ol+FIIL+FC_ZAMAN_SIMDIKI_GELECEK_Ay+EKFIIL_SAHIS_BEN_män
</S>
```

Şekil 4.4: Biçimbilimsel belirsizlik giderme için hazırlanan veri setinden bir kesit.

Bu formatta cümleler her bir kelimenin biçimbilimsel çözümleri ile MorphNet veri setine eklenmiş ve yeterli eğitimden sonra testler gerçekleştirilerek sonuçları paylaşılmıştır. Uygulama ile ilgili detaylar bir sonraki alt bölümde örnekler ile verilmiştir.

4.2.2. MorphNet ile Uygulama ve Örnekler

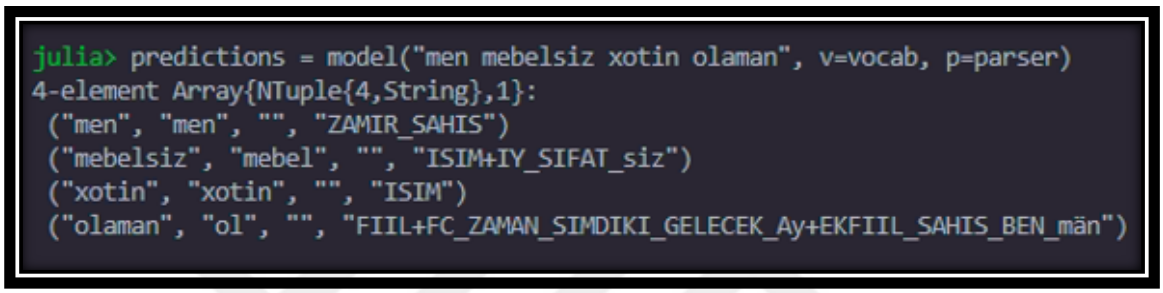
Özbekçenin biçimbilimsel belirsizlik gidermesi için Akyürek (2019) tarafından açık kaynak kodlu olarak sunulan uygulama aracı kullanılmıştır. Julia programlama dili ile kodlanan uygulamanın derlenebilmesi için GPU içeren Nvdia ekran kartına ihtiyaç bulunmaktadır. GPU desteği bulunmayan ekran kartları ile derleme yapılamamaktadır. Uygulama aşamasında 8 GB GPU’ya sahip Nvdia bir ekran kartı kullanan bir bilgisayar kullanılmıştır.

Uygulama kaynak kodları ilk indirildiğinde eğitim (train) aşasında Özbekçe için derleme yapıldığında hata alındığı için birtakım optimizasyonlar ve düzenlemeler yapılmıştır. Özbekçede arada boşluk kullanarak bir kelime tanımlanabilir. Bu gibi durumların yönetilebilmesi için düzenlemeler yapılmış derlemenin yapıldığı cihazın limitleri nedeniyle setoptim fonksiyonu üzerine yazılarak düzenlenmiştir.

Bu çalışma için i7 işlemcili, Nvidia ekran kartlı 8 GB GPU, 16 GB RAM ve 256 GB SSD içeren bir bilgisayar ile 30.000 kelime içeren bir veri seti için 8-12 saat arasında iki farklı eğitim (train) yapılmıştır. Her eğitimden (train) sonra daha doğru ve istenilen sonuçlara yakın sonuçların alındığı tespit edilmiştir. Uygulama aşamasında test için 1 dakikadan daha az yapılan eğitimde rastgele sonuçlar üretilmiştir. Ortalama 20 dakika eğitimden sonra cümlenin ilk iki kelimesi doğru olarak çözümlenmiş kalan diğer kelimeler rastgele veya varsayılan olarak bir sonuç

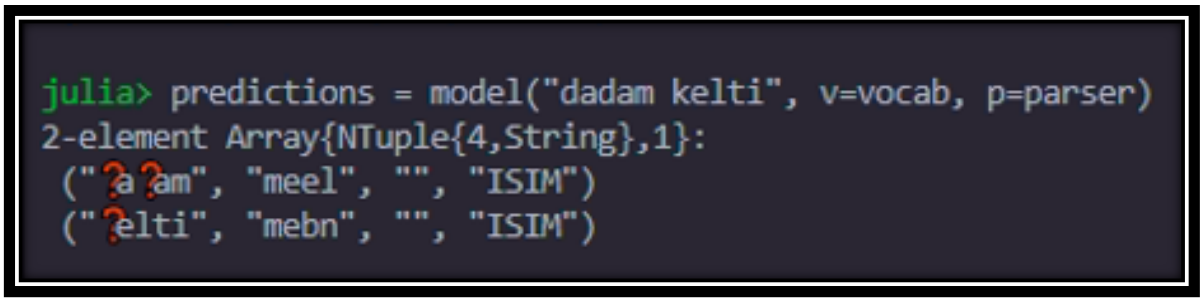
üretmemiş, sonuç bölümü boş olarak gösterilmiştir. Bu çalışmada paylaşılan örnekler ve sonuçları için ortalama 8 saat ve üzeri birden fazla eğitim yapılmıştır.

Yeterli eğitimden sonra "Ben evimi döşemeden, mobilya almadan evleneceğim, hatun alacağım" anlamına gelen "men mebelsiz xotin olaman (ben mobilyasız hatun alacağım)" cümlesinin biçimbilimsel belirsizlik giderme sonrası birden fazla çözüme sahip kelimelerin tek bir çözüme ulaştığı gözlemlenmiştir. Uygulama çıktısının görüntüsü Şekil 4.5'te paylaşılmış olup, bu cümle için %100 başarılı olarak çözümlemeyi tamamlamıştır.



Şekil 4.5: "Men mebelsiz xotin olaman" cümlesinin uygulama çıktısı.

Veri setinde yer almayan ve öğretilmeyen bir cümlenin çıktısı için sonuç üretememiştir. Şekil 4.6'da gösterildiği üzere rastgele bir sonuç üretmiştir. MorphNet modeli biçimbilimsel çözümleme ve biçimbilimsel belirsizlik giderme işlemlerini sıralı bir şekilde yapabilse de sadece çözümleme için bile yeterli miktarda eğitim (train) yapılması gerekmektedir.



Şekil 4.6: Veri seti içerisinde yer almayan "dadam kelti (babam geldi)" cümlesinin uygulama çıktısı.

MorphNet, veri setinde yer alan bir kelimenin cümle içerisinde olmayan tek kelimelik kullanımında, model öğrendiği şekilde kelimenin çözümlemesini ve tek bir çözümün sunulmasını sağlamaktadır. Aşağıdaki ekran görüntüsünde mebelsiz örneğinde bu kelimenin cümle içerisinde geçen ve öğretilen en doğru çözümünü çıktı olarak sunmaktadır.

```
julia> predictions = model("mebelsiz", v=vocab, p=parser)
1-element Array{NTuple{4,String},1}:
 ("mebelsiz", "mebel", "", "ISIM+IY_SIFAT_siz")
```

Şekil 4.7: Veri setinde cümle içerisinde yer alan bir kelimenin çözümü.

Uygulama aşamasında sağlanan veri setinin %80'i eğitim (train), %10'u geliştirme (development) ve %10'u test için kullanılmıştır. Modelin yapısı gereği bu bölümleme üç farklı parçaya rastgele olarak yapılmaktadır. Çalışmanın sonuçları sonraki bölümde tartışılmakta olup, tekniğin bilinen durumuna göre kabul edilebilir çıktılar sağlanmıştır.

4.3. NUVE PARSER İLE ÖZBEKÇENİN SÖZDİZİMSEL ÇÖZÜMLEMESİ

Tezin bir önceki bölümünde detayları ile anlatılan Özbekçenin sözdizimsel çözümlemesinin (syntactic parsing) uygulama aşamaları ve çıktıları bu bölümde detaylandırılmıştır. Zafer (2011) çalışmasında açık kaynak kodlu olarak sunulan çözümleyici geliştirilerek Nuve Parser oluşturulmuştur. Nuve Parser bu çalışma için oluşturulmuş olup, Özbekçenin sözdizimsel çözümlemesi için içerik, gramer ve istisnaların yönetimi için hem yazılımda hem de dil kurallarında düzenleme ve geliştirme yapılmıştır. Böylece Özbekçenin otomatik olarak sözdizimsel analizi yapılması öngörülmüştür. Bu çalışma için Özbekçenin biçimbilimsel çözümlemesinde ve biçimbilimsel belirsizlik giderme testlerinde de kullanılan kelimeler biçimbilimsel çözümleme ve belirsizlik giderme sonrası elde edilen veriler ile 30.000 kelime sözdizimsel çözümleme için kullanılarak, uygulama gerçekleştirilmiştir.

Nuve Parser'ın çalışma mantığı yukarıda anlatılan sözdizimsel çözümleme yöntem ve algoritmasına uygun olarak hazırlanmıştır. Buna göre, sözdizimsel çözümleme için "Yukarıdan – Aşağıya (Top – Down)" sözdizimsel çözümleyici yaklaşımına sahip, morfolojik çözümlemeye (morphological analysis), bağlamdan bağımsız gramer (Context Free Grammar) kurallarına ve geçerlilik-doğrulama kurallarına (validation rules) dayanmaktadır. Bunun için ilk olarak biçimbilimsel çözümleme için Nuve Studio ile elde edilen kelimelerin biçimbilimsel çözümlenmiş ve MorphNet ile belirsizliği giderilmiş çıktıları Nuve Parser çözümleyicisinin uygun formatına otomatik olarak dönüştürülmesi için geliştirmeler yapılarak, kelime ve kelimelerin biçimbilimsel çözümlemesinin ve sonrasında biçimbilimsel belirsizlik giderme

işlemleri ile sözdizimsel çözümleme için veri seti oluşturulmuştur. Özbekçe dil kuralları da sisteme tanımlanarak, Nuve Parser uygulamasına yazılan Özbekçe bir cümlenin Özbekçe gramer ve doğrulama kurallarına uygun bir şekilde veri setindeki tanımlamaları doğrultusunda sözdizimsel çözümlemesi otomatik olarak gerçekleştirilmiştir.

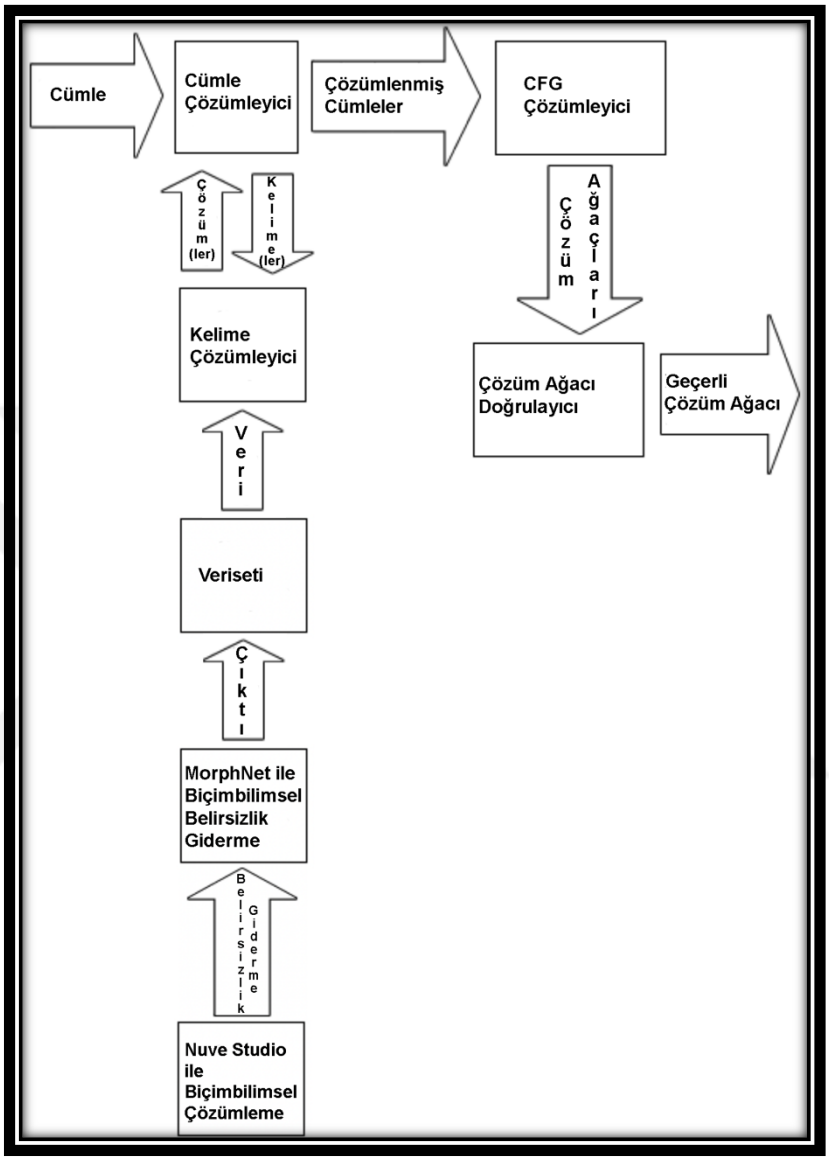
Nuve Parser ile çıktılar ve tanımlamalar daha sonra yapılacak akademik çalışmalar için Türkçe, İngilizce ve Özbekçe dillerinde dinamik üretilebilmekte olup, esnek yapısı ile daha farklı diller dinamik olarak eklenebilmektedir. Nuve Parser için sağlanacak olan ve kelime analizinde kullanılan, biçimbilimsel çözümlenmiş ve belirsizliği giderilmiş veri seti için istenilen method ve çözümleyici kullanılabilir, hatta bu çözümlemeler manuel olarak bile sunulabilmektedir. Burada bu çalışma için Nuve uygulamaları bir arada kullanılmış olsa da ayrı ayrı her bir Nuve uygulama aracının kullanılabileceği dinamik bir yapı hazırlanmıştır.

Nuve parser ile Özbekçe özelinde herhangi bir dilin sözdizimsel çözümlemesinin yapılabilmesi için;

- Dilin (Özbekçenin) gramer kurallarına (Bu kurallar temelinde sözdizimsel çözümleme sağlanmaktadır.),
- Dilin (Özbekçenin) doğrulama kurallarına (Dilin Gramer kurallarına uygun olmayan yapılar çözümleyicinin performansını artırmak için işleme tutulmadan ayıklanmaktadır, bu gramer doğrulama kuralları ile sağlanmaktadır.),
- Biçimbilimsel çözümlenmiş ve belirsizliği giderilmiş veri setine (Bu çalışma için Nuve Studio ile biçimbilimsel olarak çözümlenen veri seti daha sonra belirsizlik durumlarının giderilmesi için MorphNet ile uygulaması yapılmış ve çıktılar Nuve Parser'a eklenen bir metot ile otomatik olarak Nuve Parser veri setine uygun olarak dönüştürülmüştür.),
- Çözüm ağacı doğrulama kurallarına (Çözümleme sonrasında bulunan çözümlerin doğruluğunun, geçerli olduğunun kontrolü için bu kurallara ihtiyaç duyulmaktadır.),
- Çözümleme için ekler listesine ihtiyaç duyulmaktadır (Bu çalışmada Nuve Studio'da yer alan ekler hazırlanan bir metot ile otomatik olarak Nuve Parser için uygun biçime döndürülmüştür, bu bölüm ihtiyaca uygun olarak elle veya başka bir ek formatında hazırlanabilmektedir).

Yukarıda anlatıldığı üzere Nuve Parser'ın dinamik ve esnek olması için bir kısıtlama uygulanmamış olup, Özbekçe için yapılan çalışmada Nuve Studio ve MorphNet çıktıları

kullanılmıştır. Şekil 4.8'de bu çalışmada kullanılan yöntemlere uygun ve Özbekçe için kullanılan Nuve Parser sözdizimsel çözümleyicisinin akış şeması sunulmuştur.



Şekil 4.8: Nuve Parser sözdizimsel çözümleyici akış diyagramı.

Nuve Parser sözdizimsel çözümleyicinin akış diyagramında gösterildiği üzere, sözdizimsel çözümleme için öncelikle biçimbilimsel çözümleme çıktılarına ve bu çıktıların belirsizliklerinin giderilmesine ihtiyaç duyulmaktadır. Veri seti biçimbilimsel olarak çözümlenmiş ve belirsizliği giderilmiş kelimelerden oluşmaktadır. Veriseti hazırlanırken farklı biçimbilimsel çözümleyiciler kullanılabilir hatta elle çıktılar sağlanabilir. Önemli olan bu çıktıların Nuve Parser formatına uygun olmasıdır. Nuve Studio ve MorphNet tarafından üretilen çıktıların Nuve Parser'a uygun sözcük sınıflarına (POS) çevirimi hazırlanan metotlar ile otomatik olarak

gerçekleştirilmiştir. Böylece Nuve Parser tarafından kullanılacak veriseti hazırlanmış olmaktadır. Veriseti ile gramer kuralları ve doğrulama kuralları hazırlandıktan sonra Nuve Parser çalıştırmaya hazır hale getirilmiştir.

Nuve Parser'a sözdizimsel çözümleme için verilen bir cümle öncelikli olarak cümle çözümlemesinden geçtikten sonra her bir kelimenin çözümlenmiş hali ile çözümlenmiş cümle oluşturulmaktadır. Cümleyi oluşturan her bir kelimenin çözümlenmesi için biçimbilimsel çözümlenmiş ve belirsizliği giderilerek Nuve Parser için uygun sözcük sınıfları ile oluşturulmuş veriseti kullanılmaktadır. Cümle çözümlemesi için cümleyi oluşturan her bir kelimenin biçimbilimsel çözümlemesi kelime çözümleyici aşamasında gerçekleştirilmektedir. Biçimbilimsel çözümlemeleri yapılan ve cümleyi oluşturan her bir kelime kökleri ve ekleri ile çözümlenerek cümle tekrara oluşturulur. Bu aşama çözümlenmiş cümleler aşamasıdır ve çözümlenmiş cümleler oluşturulmaktadır. Biçimbilimsel olarak çözümlenen cümle(ler), dilin gramer kurallarına göre sözdizimsel olarak çözümlenmeye başlar. CFG çözümleyicisinin doğru bir şekilde çalışması için dilin gramer kurallarının bağlamdan bağımsız gramer kuralları ile tanımlanması ve cümle sözdiziminin bu kurallar ile doğru bir şekilde ifade edilmesi gerekmektedir. Bu kurallar çerçevesinde CFG çözümleyici cümlenin sözdizimsel çözümlerini üreterek çözüm ağaçlarını oluşturur. Nuve Parser doğru ve performanslı bir çözüm sunması açısından geçerlilik kurallarına sahiptir. Üretilen çözüm ağaçlarından hepsi belirlenen geçerlilik kurallarına göre geçerli olmayabilir. Üretilen çözüm ağaçları dilbilgisel olarak geçerli olsa da anlambilimsel açıdan geçerli olmayabilir. Örneğin, Özbekçede bir cümlede birden fazla özne bulunamaz kuralı cümlenin sözdizimsel çözümlemesi için bir geçerlilik kuralıdır. Birden fazla öznesi olan bir cümle performans açısından geçerlilik kurallarını geçemediği için işleme alınmadığı gibi geçersiz olarak sınıflandırılır. Üretilen çözüm ağaçları, çözüm ağacı doğrulayıcı aşamasında geçerlilik kurallarına göre sınıflandırılarak hem dilbilimsel hem de anlambilimsel olarak geçerli olan çözümler, geçerli çözüm ağacı aşamasında geçerli çözüm ağacı olarak sunulmaktadır.

Nuve Parser ile sözdizimsel çözümlemesi gerçekleşen bir cümlenin çıktısında aşağıdaki bilgiler yer almaktadır.

- Cümlenin varsa farklı türevleri,
- Her bir farklı türev için bulunan çözümler,

- Her bir çözüm içerisinde dilin gramerine göre hazırlanan ve bağlamdan bağımsız gramere uyarlanarak tanımlanan ve çözümü oluşturan elemanlar (elements) ve kelime grupları (word groups) sunulur.
- Elemanlar cümle (sentence=S) ve alt cümle (sub-sentence=SS) gibi kısaltmalar ile parantez ve çözüm ağacı gösteriminde sunulur.
- Kelime grupları isim tamlaması (noun complement=NC) ve sıfat tamlaması (adjective complement=JC) gibi kısaltmalar ile parantez ve çözüm ağacı gösteriminde sunulmaktadır.
- Sözdizimsel çözümlemesi tamamlanmış ve parantez ya da çözüm ağacı ile gösterilen yapıda, cümle ve alt cümleler gösterilerek, içerisinde yer alan kelimelerin ve kelime gruplarının sözcük sınıflarına (POS) göre tanımlamaları,
- Sözdizimsel olarak çözümlenmesi istenilen cümlenin hazırlanan dilin (Özbekçenin) gramer ve geçerlilik kurallarına göre geçerli olup olmadığının bilgisi sunulmaktadır. Parametrik bir ayarlama ile geçerli olmayan çözümler ayrıca sunulmaktadır.
- Parantez gösterimi ile sunulan çözümün çözümleme ağacı ve geçerlilik bilgileri yer almaktadır.

Özbekçenin sözdizimsel çözümlemesinin Nuve Parser ile uygulanması için detayları alt bölümlerde anlatılan hazırlıklar yapılarak ilgili tanımlama dosyaları oluşturulmuştur ve yazılım üzerinde geliştirmeler sağlanarak bazı dosyaların otomatik olarak hazırlanması sağlanmıştır. Çıktının istenilen formatta oluşturulması için hem Nuve Parser hem de gelecekte yapılması planlanan geliştirmeler için Nuve Studio üzerinde bazı geliştirmeler yapılmıştır. Sonraki alt bölümlerde uygulama aşamaları, hazırlanan Nuve Parser dosyaları, geliştirmeler ve çıktıları detayları ile verilmiş olup, Özbekçenin sözdizimsel çözümlemesinin Nuve Parser ile uygulaması dört kısımda detaylandırılmıştır.

4.3.1. Özbekçe Gramer Kurallarının Hazırlanması

Nuve Parser ile Özbekçenin sözdizimsel çözümlemesinin yapılması için öncelikle Özbekçe gramer kurallarının, bağlamdan bağımsız gramer kuralları ile uygulama altyapısına uygun bir şekilde XML biçiminde hazırlanması gerekmektedir. Bu dosya (grammar.xml) içerisinde Özbekçeye ait sözdizimsel çözümleme için kullanılacak tüm kurallar tanımlanmalıdır. Aksi takdirde tanımlanmayan bir kural nedeniyle istenilen çözüm sağlanamayacaktır.

Özbekçe gramer kurallarının oluşturulması için cümleyi oluşturan her bir bileşen (özne, nesne, yüklem, isim tamlaması vb.) için kuralların yazılması gerekir. Özbekçe gramer kuralları bağlamdan bağımsız gramer olarak grammar.xml dosyası içerisinde ifade edilmektedir. Her bir kural rule tagları içerisine yazılmakta olup, category tagı içerisinde kuralı oluşturan bileşenler isimleri, sonlu veya devamlı oldukları gibi bağlamdan bağımsız gramer ile ifade edilebilecek bileşenler xml olarak ifade edilmektedir. Aşağıda bağlamdan bağımsız gramer gösteriminin, Nuve Parserda kullanılan XML gösterimi örnek olarak verilmiştir. Örnekte a ve b sonuncu oldukları için category tagına sonuncu olduklarını belirten terminal niteliği (attribute) eklenmiştir.

S → a | b | aSb

```
<rule category=”S”>
<category name=”aSb”/>
</rule>
<rule category=”S”>
<category terminal=”true” name=”a”/>
</rule>
<rule category=”S”>
<category terminal=”true” name=”b”/>
</rule>
```

Özbekçe gramer kurallarının yukarıdaki örnekte olduğu gibi grammar.xml dosyasında tanıtılması ve tüm ihtimallerin kural olarak eklenmesi gerekmektedir. Şekil 4.9’daki grammar.xml dosyasından sunulan bir kesitte Özbekçede yüklemi oluşturan durumlar kuralları ile örnek olarak sunulmuştur. Hazırlanan xml dosyasında, gramer adı Özbekçe olan ve yüklem olabilecek durumlar gösterilmiştir. Bu örnek kesitte gramer dosyası cümle içerisindeki yüklem durumlarını sunarak gramer dosyası kapatılmıştır. Eklenecek her bir cümle bileşeni tüm durumları (sonlu, devamlı vb.) ile kural olarak tanıtılmalıdır.

```
<?xml version="1.0"?>
<grammar name="uzbek">

<!-- START : Constituents - Ögeler-->

   <!-- START  : Predicate - Yüklem -->
   <rule category="Pre">
      <category name="VP"/>
   </rule>
   <rule category="Pre">
      <category name="VP"/>
      <category name="Que"/>
   </rule>
   <rule category="Pre">
      <category name="NP"/>
      <category name="Que"/>
   </rule>
   <rule category="Pre">
      <category name="NP"/>
      <category name="Prs"/>
       <category name="Que"/>
   </rule>
   <rule category="Pre">
      <category terminal="true" name="pre"/>
   </rule>
   <rule category="Pre">
      <category terminal="true" name="pre"/>
      <category name="Que"/>
   </rule>
   <rule category="Pre">
      <category name="NP"/>
      <category terminal="true" name="aux"/>
   </rule>
    <rule category="Pre">
      <category name="NP"/>
      <category terminal="true" name="dir"/>
   </rule>
   <rule category="Pre">
      <category name="AP"/>
      <category terminal="true" name="aux"/>
   </rule>
   <rule category="Pre">
      <category name="NP"/>
      <category name="Prs"/>
   </rule>
   <rule category="Pre">
      <category name="NP"/>
      <category name="Prs"/>
      <category terminal="true" name="aux"/>
   </rule>
   <rule category="Pre">
      <category name="NP"/>
      <category terminal="true" name="aux"/>
      <category name="Prs"/>
   </rule>
   <!-- END    : Predicate - Yüklem -->
</grammar>
```

Şekil 4.9: Özbekçede yüklemi oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit.

Yukarıdaki şekilde gösterildiği üzere, Özbekçede yüklem fiil unsurundan (VP), fiil unsuru (VP) ve soru ekinden (Que), isim unsuru (NP) ve soru ekinden (Que), isim unsuru (NP)-şahıs eki (Prs) ve soru eki (Que) gibi farklı bileşenlerden oluşabilmektedir. Bu bileşenler örnek bir kesit olarak sunulmuş olup, örnekler çoğaltılabilir.

Özbekçenin sözdizimsel çözümlemesinin bağlamdan bağımsız gramer ile sonlu durumlar ile ifade edilmesi için grammar.xml dosyası içerisinde tanımlanan tüm durumların kurallarının ve sonlu durumlarının tanımlanması gerekmektedir. Yukarıdaki yüklem örneğinde fiil unsurlarının (VP), isim unsurlarının (NP) ve soru ekininde kural tanımlamalarının ve sonlu durumlarının grammar.xml dosyası içerisinde tanımlanması gerekmektedir. Böylece Özbekçenin dil kuralları bağlamdan bağımsız gramer kuralları ile ifade edilerek sözdizimsel çözümleme gerçekleştirilecektir.

Fiil unsurlarının (VP) kuralları ve sonlu durumları grammar.xml dosyasında tanımlanmıştır. Bu tanımlamaların bir kesiti Şekil 4.10'da sunulmuş olup, örnek olarak Özbekçede fiil unsurları fiilden, fiil unsuruna eklenen olumsuzluk ve zaman eklerinden, fiil-zaman-ikinci tekil iyelik eki ve soru ekinin birleşiminden, isimlerden sonra gelen -lan eki gibi ekler ile oluşturulabilir. Bu bileşenler örnek bir kesit olarak sunulmuş olup, örnekler çoğaltılabilir.

Fiil unsurları (VP) gibi yüklemi oluşturan bir diğer bileşen isim unsurlarıdır (NP). İsim unsurlarının (NP) kuralları ve sonlu durumları grammar.xml dosyasında tanımlanmıştır. Bu tanımlamalardan örnek bir kesit Şekil 4.11'de sunulmuştur. Örnek olarak Özbekçede isim unsurları isimden, zamirden, özel isimden, sıfattan, isim ve sıfata eklenen çoğul ekleri gibi ekler ile oluşturulabilir. Bu bileşenler örnek bir kesit olarak sunulmuş olup, örnekler çoğaltılabilir.

Yüklemi oluşturan örnek bileşenlerden bir diğeri soru (Que) eki unsurudur. Soru eki unsurlarının (Que) kuralları ve sonlu durumları grammar.xml dosyasında tanımlanmıştır. Bu tanımlamaların bir kesiti Şekil 4.12'de sunulmuş olup, örnek olarak Özbekçede soru eki unsurları soru ekinden, soru eki unsuruna eklenen yardımcı fiilden, soru eki unsuruna eklenen şahıs eki unsurundan, isim unsuru-şahıs eki unsuru ve soru eki unsurlarının birleşimi gibi ekler ile oluşturulabilir. Bu bileşenler örnek bir kesit olarak sunulmuş olup, örnekler çoğaltılabilir.

```
<!-- START  : Verb Phrase - Fiil Unsuru -->
  <rule category="VP">
     <category terminal="true" name="verb"/>
  </rule>
  <rule category="VP">
      <category terminal="true" name="verb"/>
     <category terminal="true" name="tense"/>
      <category terminal="true" name="p2s"/>
       <category terminal="true" name="que"/>
  </rule>
  <rule category="VP">
     <category name="VP"/>
     <category name="Prs"/>
  </rule>
  <rule category="VP">
     <category name="VP"/>
     <category terminal="true" name="aux"/>
  </rule>
  <rule category="VP">
     <category name="VP"/>
     <category terminal="true" name="mood"/>
  </rule>
   <rule category="VP">
     <category terminal="true" name="VP"/>
     <category terminal="true" name="imp"/>
  </rule>
  <rule category="VP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="lan"/>
  </rule>
<rule category="VP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="Ga"/>
  </rule>
  <rule category="VP">
     <category terminal="true" name="verb"/>
     <category terminal="true" name="psv"/>
  </rule>
  <rule category="VP">
     <category name="VP"/>
     <category terminal="true" name="neg"/>
  </rule>
  <rule category="VP">
     <category name="VP"/>
     <category terminal="true" name="tense"/>
  </rule>
<!-- END    : Verb Phrase - Fiil Unsuru -->
```

Şekil 4.10: Özbekçede fiil unsurlarını oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit.

```
<!-- START  : Noun Phrase - İsim Unsuru -->
  <rule category="NP">
     <category terminal="true" name="noun"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="pron"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="prop"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="adj"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="plu"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="adj"/>
     <category terminal="true" name="plu"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="JC"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="prop"/>
     <category terminal="true" name="plu"/>
  </rule>
  <rule category="NP">
     <category name="NP"/>
     <category terminal="true" name="gen"/>
     <category terminal="true" name="rel"/>
  </rule>
   <rule category="NP">
     <category terminal="true" name="adj"/>
     <category terminal="true" name="lik"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="li"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="ta"/>
  </rule>
  <rule category="NP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="siz"/>
  </rule>
  <rule category="NP">
     <category name="NP"/>
     <category name="Poss"/>
  </rule>
     <rule category="NP">
     <category terminal="true" name="noun"/>
     <category terminal="true" name="Ga"/>
  </rule>
<!-- END    : Noun Phrase - İsim Unsuru -->
```

Şekil 4.11: Özbekçede isim unsurlarını oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit.

```
<!-- START  : Question - Soru -->
   <rule category="Que">
      <category terminal="true" name="que"/>
   </rule>
   <rule category="Que">
      <category name="Que"/>
      <category terminal="true" name="aux"/>
   </rule>
   <rule category="Que">
      <category name="Que"/>
      <category name="Prs"/>
   </rule>
    <rule category="Que">
      <category name="NP"/>
      <category name="Prs"/>
      <category name="Que"/>
   </rule>
<!-- END    : Question - Soru -->
```

Şekil 4.12: Özbekçede soru eki unsurlarını oluşturan bileşenlerin gösterildiği ve kurallarının tanımlandığı grammar.xml dosyasından bir kesit.

Bu bölümde örnek bileşenler için sunulan kurallar cümle içerisinde yer alan tüm bileşenler için hazırlanmıştır. Nuve Parser ile sözdizimsel çözümlemenin başarı oranı tüm bileşenler için tanımlanan kuralların bütün ihtimalleri ile tanımlanmasına bağlı olup, bu kurallar grammar.xml dosyasında yer almalıdır. Özbekçenin sözdizimsel çözümlemesi için cümleyi oluşturan bileşenlerin mümkün olan tüm ihtimalleri ile eklenerek kuralların tanımlanması sağlanmıştır.

4.3.2. Nuve Parser için Özbekçe Veri Setinin Hazırlanması

Özbekçenin sözdizimsel çözümlemesinin Nuve Parser ile yapılabilmesi için veri setinde kullanılacak her bir kelimenin sırasıyla biçimbilimsel çözümlemesinin yapılması, sonrasında her bir kelimenin biçimbilimsel belirsiz durumlarının giderilmesi ve Nuve Parser için uygun olan tanımlı bileşen etiketlerinin (tag) kullanılması gerekmektedir.

Bu çalışmada kullanılan Nuve Studio, MorphNet ve Nuve Parser açık kaynak kodlu yapıların geliştirilmesi ile bu çalışma için hazırlanmış olsa da her bir uygulamanın bağımsız kullanılması planlandığı için tüm uygulamalar kendilerine özgü veri modeline veya işleyişe sahip olup, bu veri modelleri birbirlerine otomatik olarak dönüştürülecek veya ortak kullanıma uygun bir şekilde hazırlanmıştır.

Nuve Parser tarafında birçok geliştirme yapılmıştır. Ancak bu geliştirmeler Nuve Parser'ın esnek ve bağımsız kullanımına engel olmayacak şekilde planlanmıştır. Nuve Parser için hazırlanan veri seti elle oluşturulmuş olabilir ve bu durum Nuve Parserın çalışmasını engellememektedir.

Özbekçenin sözdizimsel çözümlemesi için hazırlanan veri seti için biçimbilimsel çözümleme Nuve Studio biçimbilimsel çözümleme aracı ile yapılmıştır. Biçimbilimsel çözümleme sonrası ortaya çıkan birden fazla çözümün yer alması ve belirsiz durumlar için biçimbilimsel belirsizlik giderme MorphNet uygulaması ile yapılmıştır. Biçimbilimsel belirsizliği giderilen kelimelerin çözümlerinin Nuve Parser bileşen etiketlerine uygun olarak çevirimi sağlanarak Nuve Parser ile sözdizimsel çözümleme sağlanmıştır. Örnek olarak Shirin nega uxlamayapsan (Shirin neden uyumuyorsun)? soru cümlesini verisetine eklemek istediğimizde öncelikli olarak bu cümleyi Nuve Studio ile biçimbilimsel çözümlemesini gerçekleştirerek aşağıdaki çıktılar alınmaktadır.

Shirin

Shirin/ISIM_OZEL

nega

ne/ZAMIR_SORU IC_HAL_YONELME_Ga

nega/ISIM

ne/ISIM IY_ISIM_AQ IY_ISIM_ä

ne/ISIM IY_ISIM_AQ IY_FIIL_ä

ne/ISIM IY_ISIM_Ga

ne/ISIM IY_FIIL_Ga

uxlamayapsan

uxla/FIIL OLUMSUZLUK_mä FC_ZAMAN_SIMDIKI_yäp EKFIIL_SAHIS_SEN_sän

uxla/FIIL FY_ISIM_Im IY_FIIL_ä FC_ZAMAN_SIMDIKI_yäp EKFIIL_SAHIS_SEN_sän

uxla/FIIL FY_FIIL_mä FC_ZAMAN_SIMDIKI_yäp EKFIIL_SAHIS_SEN_sän

Nuve Studio biçimbilimsel çözümleyicisi çözümlenecek kelimenin gövdesini ve türünü "/" işareti ile gösterdikten sonra gelen her bir ekin türünü birer boşluk bırakarak gövdeye eklemektedir. Ayrıca bu çözümde nega ve uxlamayapsan kelimelerinin birden fazla biçimbilimsel çözümü üretilmiş olup biçimbilimsel belirsizlik giderme sonrasında Nuve Parser veri setine eklenebilecektir.

Biçimbilimsel belirsizlik giderme için MorphNet uygulaması kullanılmış olup bu çıktı biçimi ile MorphNet veriseti formatı farklılık göstermektedir. Bu nedenle hazırlanan bir script ile Nuve Studio çıktıları MorphNet veriseti formatına çevirilerek aşağıdaki veri girişi MorphNet uygulamasına sağlanmıştır. Aşağıda biçimbilimsel çözümlemesi Nuve Studio ile tanımlanan ve MorphNet veriseti formatına uyarlanan cümlenin örneği gösterilmektedir.

<S>

Shirin Shirin+ISIM_OZEL

nega ne+ZAMIR_SORU+IC_HAL_YONELME_Ga nega+ISIM
ne+ISIM+IY_ISIM_AQ+IY_ISIM_ä ne+ISIM+IY_ISIM_AQ+IY_FIIL_ä
ne+ISIM+IY_ISIM_Ga ne+ISIM+IY_FIIL_GA

uxlamayapsan

uxla+FIIL+OLUMSUZLUK_mä+FC_ZAMAN_SIMDIKI_yäp+EKFIIL_SAHIS_SEN_sän
uxla+FIIL+FY_ISIM_Im+IY_FIIL_ä+FC_ZAMAN_SIMDIKI_yäp

+EKFIIL_SAHIS_SEN_sän

uxla+FIIL+FY_FIIL_mä+FC_ZAMAN_SIMDIKI_yäp+EKFIIL_SAHIS_SEN_sän

</S>

MorphNet uygulamasının veriseti formatında eklenecek cümle açılıp kapatılan S etiketleri (tag) arasında tanımlanarak, her bir kelimenin biçimbilimsel çözümleri bir tab karakter boşluğunda eklenmeli ve kelimenin ilk biçimbilimsel çözümü, elle o cümle için biçimbilimsel olarak doğru (istenilen) çözümlenmiş yani belirsizliği giderilmiş çözümü olmalıdır. Kelime gövdesine

eklenen ekler + işareti ile ayrılmaktadır. MorphNet veriseti istenilen formatta hazırlandıktan sonra yeterli veri ve eğitimden sonra uygulamanın çıktısı aşağıdaki gibi olmaktadır.

(“Shirin”,”shirin”,””,”ISIM_OZEL”)

(“nega”,”ne”,””,”ZAMIR_SORU+IC_HAL_YONELME_Ga”)

(“uxlamayapsan”,”uxla”,””,”
FIIL+OLUMSUZLUK_mä+FC_ZAMAN_SIMDIKI_yäp+EKFIIL_SAHIS_SEN_sän”)

Biçimbilimsel belirsizlik giderme sonrası her bir kelimenin tek bir biçimbilimsel çözümü gösterilmekte olup, sözdizimsel çözümlemede her bir kelimenin bir biçimbilimsel çözümünün olması gerekmektedir. Son olarak veri setinin sözdizimsel çözümlemede kullanılabilmesi için Nuve Parser veri seti formatına uygun hale getirilmelidir. Bu uygulamalarda geliştirmeler yapılırken her bir uygulamanın (Nuve Studio, Nuve Parser vb.), ayrı ayrı birbirinden bağımsız kullanılabilir olması planlanmıştır. Bu nedenle veri seti formatları farklı olsa da çevirim için bir katman, script ya da metot sağlanarak uygulamalar arasında geçiş kolaylıkla sağlanmıştır ve ortak bir veri setinin kullanımına imkân tanınmıştır. Böylece veri setinin gelişimi her bir uygulamanın gelişimini sağlamıştır. Nuve Parser sözdizimsel çözümlemesinde kullanılacak eklerin isimlendirilmesi Nuve Studio ve MorphNet uygulamalarından farklıdır. Bunun için bir çevirim tablosu kullanılarak her bir ekin Nuve Parser karşılığı (Tag) uygulama içerisinde yer alan manuel_tags dosyası ile sağlanmıştır. Buna göre Nuve Studio ekler dosyası otomatik olarak işlenerek ilgili ek karşılığı (tag) manuel_tags dosyasına eklenmiş ve bu dosya üzerinden çevirimler sağlanmıştır. Çevirim tablosundan örnek bir kesit Tablo 4.3’te gösterilmiştir. Ayrıca Nuve Parser veri setinde tanımlanan ek isimlerinin yanı sıra veri formatı da farklılık göstermekte olup, Nuve Parser kelime çözümleyici her bir gövde ve ekten sonra gövdenin veya ekin türünü gösteren bir formatı kabul edecek şekilde geliştirilmiştir. Buna göre örnek cümlemizin Nuve Parser veri setindeki formatı aşağıdaki gibidir.

Shirin Shirin/prop

nega ne/pron+ga/dat

uxlamayapsan uxla/verb+ma/neg+yap/tense+san/2sg

Tablo 4.3: Tag ve eklerin eşleştirilmesini sağlayan çevirim tablosundan bir kesit.

Ek	Tag
ISIM	noun
ISIM_OZEL	prop
SIFAT	adj
ZAMAN	adv
ZAMIR	pron
ZAMIR_SAHIS	pron
FIIL	verb
EDAT	posp
SAYI	num
BAGLAC	conj
SORU	que
ISIM_COGUL_LER	plu
ISIM_TAMLAMA_IN	gen
ZAMIR_SAHIPLIK_IM	gen
ISIM_YONELME_E	dat
ISIM_KALMA_DE	loc
ISIM_CIKMA_DEN	abl
ISIM_BELIRTME_I	acc
ISIM_BIRLIKTELIK_LE	ins
ISIM_BULUNMA_KI	rel

Yukarıda çevirim tablosundan bir kesit sunulmuş olup, Nuve Parser verisetinin oluşturulması için tüm eklerin karşılığı bu tabloda tanımlanmış ve Nuve Studio ekler dosyasından alınan tüm eklerin karşılıkları bu tablo içerisinde manuel_tags dosyasında tanımlanmıştır.

4.3.3. Nuve Parser için Özbekçe Doğrulama Kurallarının Hazırlanması

Bağlamdan bağımsız gramer kullanılarak doğal dillerin sözdizimsel çözümlemesi karmaşık olmasının yanında en uygun yöntem de değildir. Bunun için bir çok kontrol kuralı tanımlanmasına rağmen bazı net durumlarda bir çok kural yazılmadan geçersiz durumlar nedeniyle sonucun geçersiz olduğunu göstermektedir. Böylece hem uygulamanın performansı hem de birçok kuralın tanımlanmasına gerek kalmadan sonuca erişilmesine imkân tanımıştır.

Eklenen bu ek modül sayesinde çözüm ağaçlarından geçerli olanlar tespit edilerek, geçerli çözümlerin çıktıları uygulama tarafından sağlanmaktadır. Nuve Parser tarafında bu uygulama özelliği parametrik olarak ayarlanmış olup istenildiğinde geçersiz durumlarda listelenebilmektedir.

Çözüm ağacı doğrulayıcı (Parse tree validator) gramer dosyasına benzer şekilde kurallardan oluşmaktadır. Bu kurallar validation_rule.xml dosyasının içerisinde rule taglarının içerisinde tanımlanarak cümle içerisinde olması ya da olmaması gereken durumları ve kısıtlamaları kurallar ile göstermektedir. Özbekçe için yazılan bu doğrulama kuralları birçok gramer kuralının yazılmadan aynı seviyede kontrole olanak sağlamaktadır. Bu kurallar cümle yapılarına göre, öğe sayısı kuralları, yüklem ve nesne uyumu kuralları ve tamlama kuralları gibi kurallardan oluşmaktadır. Örneğin, Özbekçede bir cümlede birden fazla özne olamaz. Bu kural aşağıdaki gibi tanımlanmaktadır.

```
<rule scope="SS" recursive="false" ignore="VJG, VNG, VAG">
    <description>Birden fazla özne olamaz</description>
    <status type="rest" cat="SS" op="notMoreThanOne" str="[Sub["/>
</rule>
```

Özbekçede yüklem ve nesne arasında bir uyum söz konusudur. Örneğin belirtili ve belirtisiz nesne bir cümlede bir arada olamaz. Böyle bir durumun olmadığının kontrolü grammar.xml

dosyasında tanımlanan gramerde olması gereken durumların tanımlamalarının yanında pek kolay tanımlanamayacağı için bu ek modüle ve kurallar dosyasına ihtiyaç duyulmuştur. Bu doğrultuda geliştirmeler ile tanımlamaları yapılmıştır. Özbekçede belirtili ve belirtisiz nesne bir arada olamaz kuralının Nuve Parserda çözüm ağacı doğrulayıcı için yazılmış kuralı aşağıdaki gibidir.

```
<rule scope="SS" ignore="VJG, VNG, VAG">
    <description>Belirtili ve Belirtisiz nesne bir arada olamaz</description>
    <status type="cond" cat="IObj" op="exists"/>
    <status type="rest" cat="DObj"  op="notExists"/>
</rule>
```

Son olarak başka bir doğrulama(geçerlilik) kuralı tamlamalar ile ilgili yazılmış olup, Özbekçede iki sıfat ile sıfat tamlaması yapılamaz ifadesi validation_rule.xml dosyasında aşağıdaki gibi ifade edilmiştir.

```
<rule scope="JC" recursive="true">
    <description>İki sıfat ile sıfat tamlaması olamaz</description>
    <status type="rest" cat="NP" op="notIncludes"  str="[NP[adj["/>
</rule>
```

Tanımlanan doğrulama (geçerlilik) kuralları parantez gösterimde ilgili kataın(string) kontrolü ile anlaşılmakta olup, kuralda tanımlanan durumun(statüs) varlığına ya da yokluğuna göre geçerli ya da geçersiz olduğu gösterilmektedir.

4.3.4. Nuve Parser ile Özbekçe Sözdizimsel Çözümlemesinin Çıktıları

Nuve Parser ile Özbekçenin sözdizimsel çözümlemesinin çıktıları parantez ve çözüm ağacı gösterimi ile sunulmaktadır. Cümlenin kaç farklı türevinin olduğu ve her bir türevin kaç çözümünün olduğu gösterilmektedir. Gelecekte Nuve Studio ile entegrasyonu planlandığı için çıktılar web ortamına uyumlu bir şekilde tasarlanmış olup sözdizimsel ağaç üreticisi (syntax

tree generator) web üzerinde çıktıları göstermektedir. Örneğin “men tushundim (ben düşündüm)” cümlesinin 1 farklı türevi bulunmuş olup, 1. türev için 1 çözüm sunulmuştur. Nuve Parser örnek çıktısı aşağıdaki şekilde sunulmuştur.



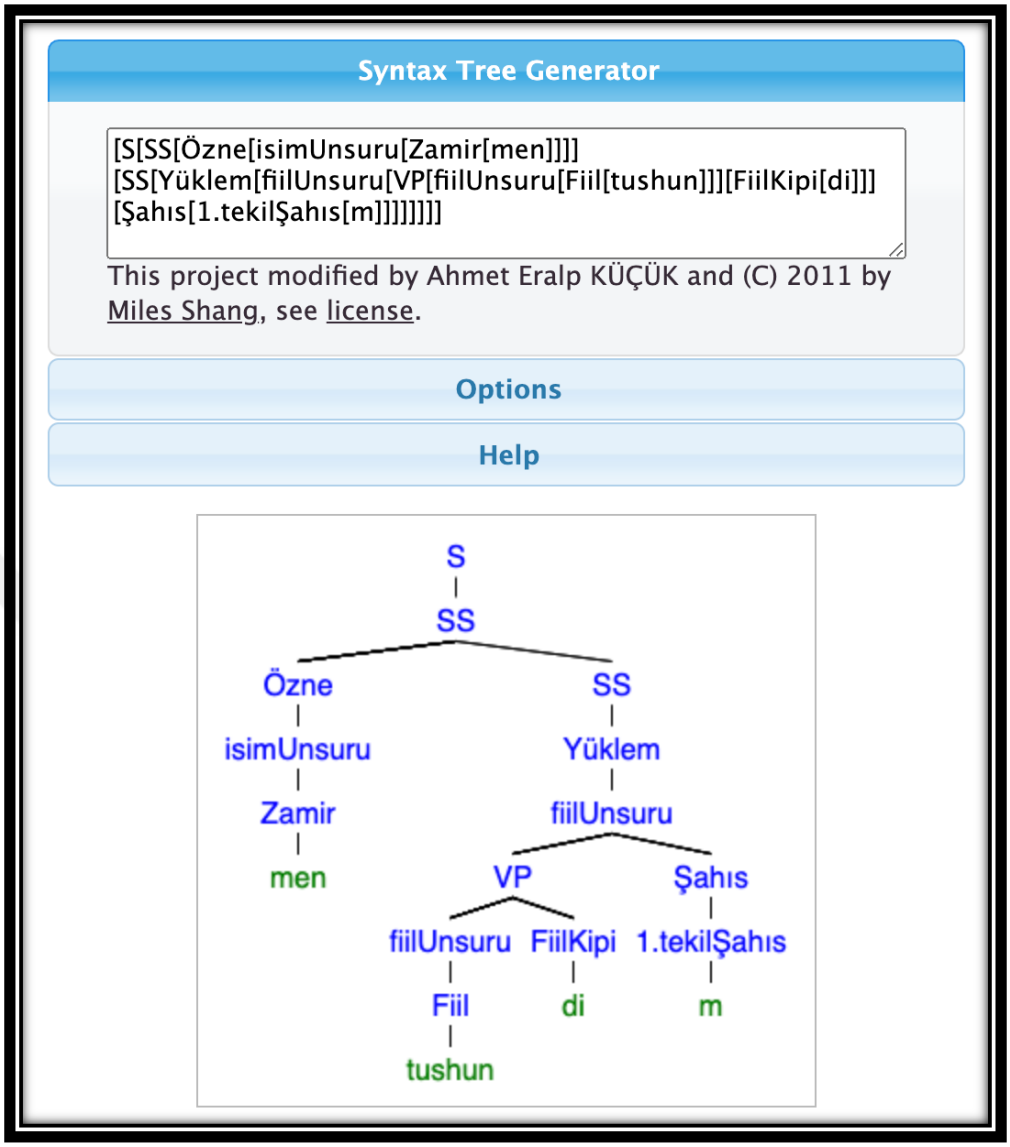
Cümlenin 1 farklı türevi bulundu.

1. türev için 1 çözüm bulundu.

"men/pron" "tushun/verb di/tense m/1sg"	Geçerli Mi?	Ağacı Çiz
[S[SS[Özne[isimUnsuru[Zamir[men]]]][SS[Yüklem[fiilUnsuru[VP[fiilUnsuru[Fiil[tushun]]][FiilKipi[di]]][Şahıs[1.tekilŞahıs[m]]]]]]]	Evet	Ağacı Çiz

Şekil 4.13: Örnek cümlenin Nuve Parser çıktısı.

Şekil 4.13’te gösterildiği üzere cümlenin içinde yer alan kelimelerin biçimbilimsel çözümlemeleri, cümlenin parantez gösterimi, geçerlilik durumu sunulmuştur. Sadece geçerli çözümler sunulmakta olup, parametrik bir ayarlama ile geçersiz çözümler ölçümlemeler için ayrıca gösterilebilmektedir. Ağacı çiz linki ile parantez gösterimi sunulan çözümün, çözüm ağacı ile gösterimi web tabanlı olarak sunulmaktadır. Linkte tıklandığında bir sonraki sayfaya veri parantez gösterimi URL üzerinden erişilebilir bir format olan GET metodu aracılığıyla iletilmektedir. Örnek cümlenin çözüm ağacının gösterildiği link aşağıdaki gibidir. nuvestudio.com/syntree/index.php?[S[SS[Özne[isimUnsuru[Zamir[men]]]][SS[Yüklem[fiilUnsuru[VP[fiilUnsuru[Fiil[tushun]]][FiilKipi[di]]][Şahıs[1.tekilŞahıs[m]]]]]]] bu link web tabanlı bir üretici (generator) tarafından çözüm ağacına dönüştürülmekte olup, aşağıdaki şekilde gösterilmiştir. Çözümleme ağacı çizim kütüphanesi geliştirme ve güncellemeler kapsamında Nuve Parser’a eklenmiş olup, Shang (2011) referans çalışması düzenlenmiş ve eklenmiştir.



Şekil 4.14: Nuve Parser web tabanlı sözdizimsel ağaç üretici.

Çıktıların web tabanlı olarak planlanmasının nedeni Nuve Studio ve Nuve Parser'ın birleştirilme düşüncesidir. Sonraki versiyonlarda bu güncelleme yapılarak Nuve Parser'ın tamamen web tabanlı ve kullanıcılar tarafından daha rahat erişilebilmesine olanak sağlaması planlanmaktadır.

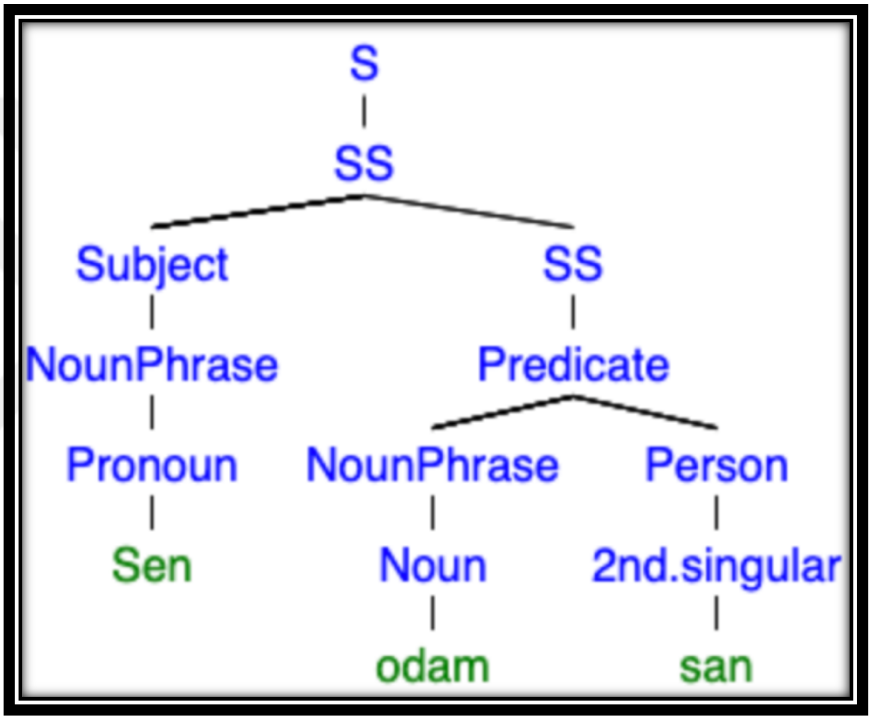
4.3.5. Nuve Parser ile Özbekçe Sözdizimsel Çözümleme Örnekleri

Bu bölümde, Nuve Parser ile Özbekçe sözdizimsel çözümleme örnekleri yapılmış ve örnek cümleler çözüm ağaçları ile sunulmuştur. Örneklerin oluşturulmasında ve sıralamasında Zafer (2011) ile Karahan (2012) çalışmaları referans alınarak, Türkçe için ölçümlenen sonuçlar ile

karşılaştırma fırsatı edinilmiştir. Buna göre, Özbekçe cümle türlerine göre örnekler aşağıdaki gibi sunulmuştur.

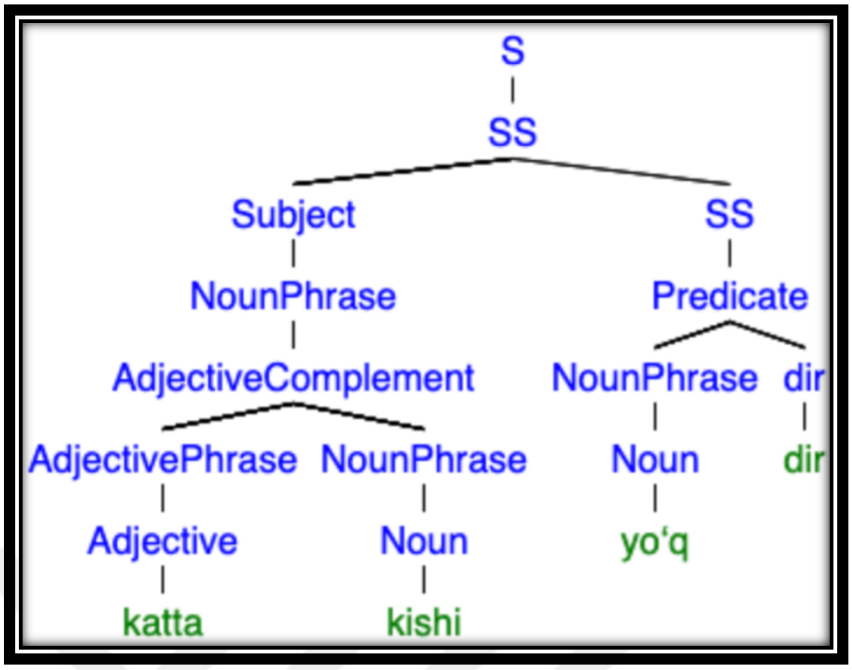
İsim cümlesi, yüklemi fiilden meydana gelmeyen isim soylu kelimelerden oluşan cümle türleridir. Aşağıda olumlu, olumsuz ve soru cümlelerine ait örnekler ve çözüm ağaçları sunulmuştur.

- Sen odamsın (Sen insansın). Olumlu isim cümlesi örneğinin çözüm ağacı Şekil 4.15'te gösterilmiştir.



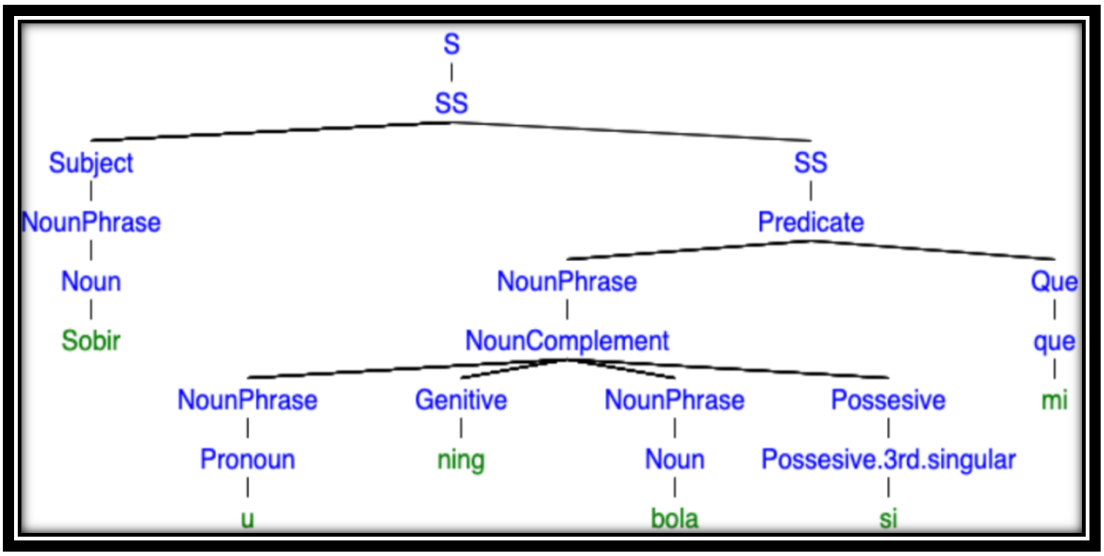
Şekil 4.15: Örnek olumlu isim cümlesi çözüm ağacı.

- Katta kishi yo'qdir (Büyük kişi yoktur). Olumsuz isim cümle örneğinin çözüm ağacı Şekil 4.16'da gösterilmiştir.



Şekil 4.16: Örnek olumsuz isim cümlesi çözüm ağacı.

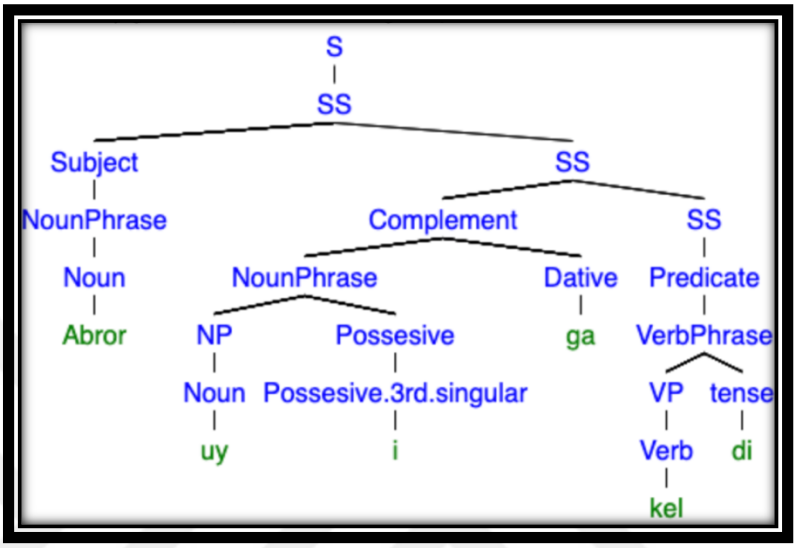
- Özbekçede soru eki bitişik olarak yazılmaktadır. Bu durum Türkçeden farklı istisnai bir durumdur. Sobir uning bolasimi (Sobir onun çocuğu mu)? Soru isim cümlesi örneğinin çözüm ağacı Şekil 4.17'de gösterilmiştir.



Şekil 4.17: Örnek soru isim cümlesi çözüm ağacı.

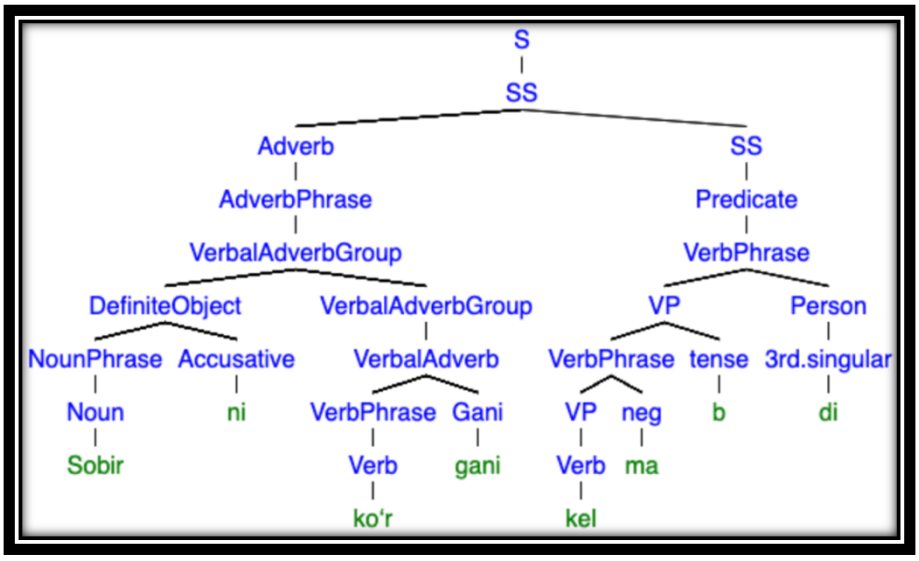
Yüklemi fiil olan cümlelere fiil cümlesi denilmektedir. Olumlu, olumsuz ve soru fiil cümlelerinin örnekleri çözüm ağaçları ile aşağıda verilmiştir.

- Abror uyiga keldi (Abror evine geldi). Olumlu fiil cümlesinin çözüm ağacı ile gösterimi aşağıda gösterilmiştir.



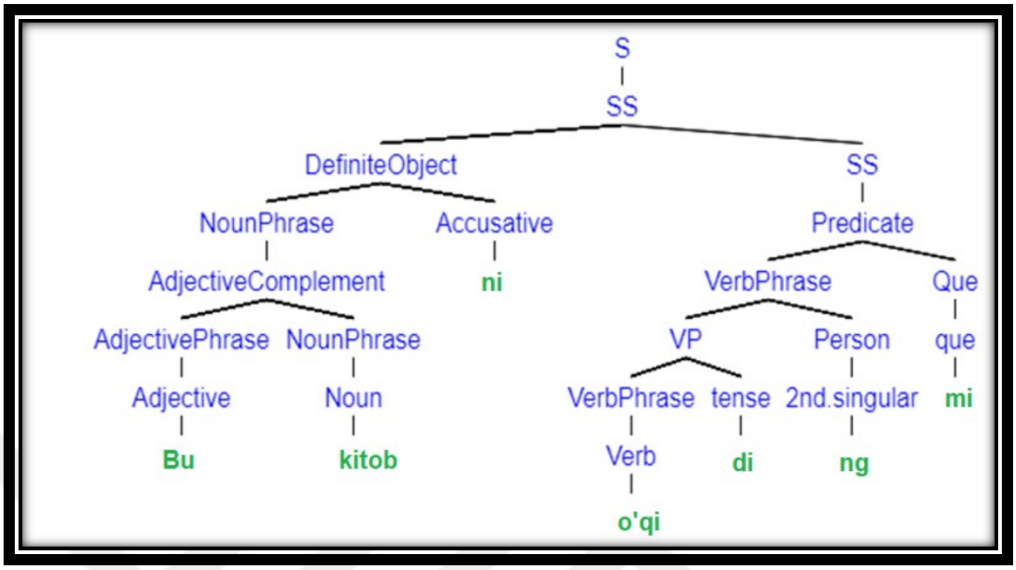
Şekil 4.18: Örnek olumlu fiil cümlesi çözüm ağacı.

- Sobirni ko'rgani kelmabdi (Sobir'i görmeye gelmemiş). Olumsuz fiil cümlesi örneğinin çözüm ağacı ile gösterimi Şekil 4.19'da gösterilmiştir.



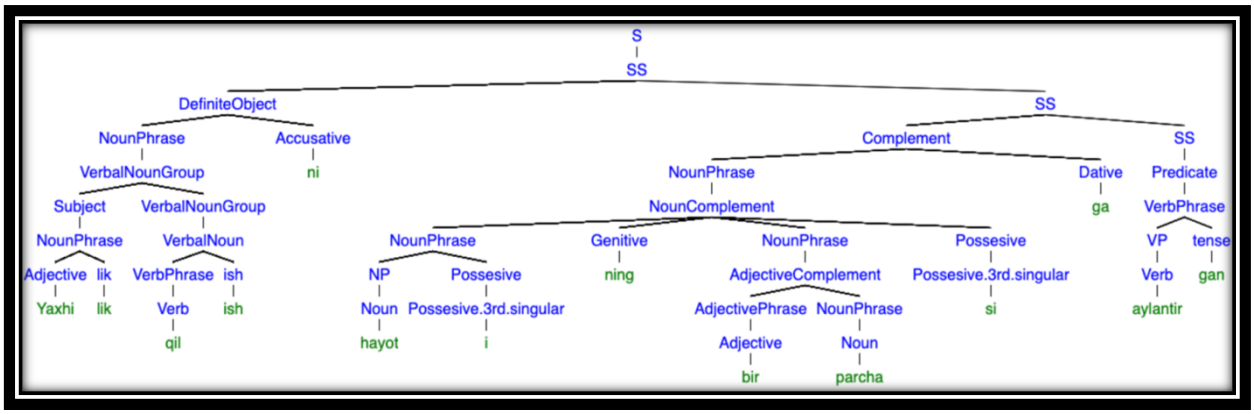
Şekil 4.19: Örnek olumsuz fiil cümlesi çözüm ağacı.

- Özbekçede soru eki bitişik olarak yazılmaktadır. Bu durum Türkçeden farklı istisnai bir durumdur. Bu kitobni o'qidingmi (Bu kitabı okudun mu)? Soru fiil cümlesi örneğinin çözüm ağacı Şekil 4.20'de gösterilmiştir.



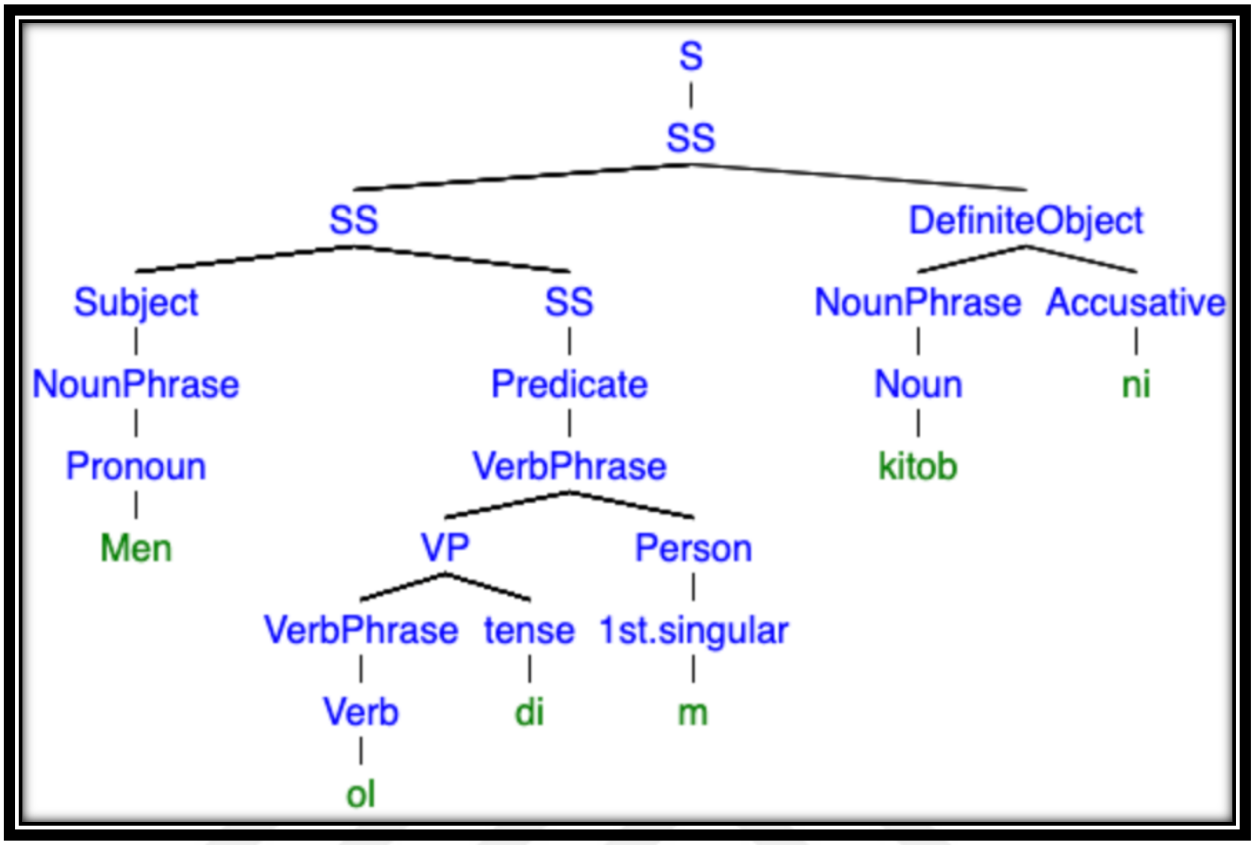
Şekil 4.20: Örnek soru fiil cümlesi çözüm ağacı.

Cümlenin öğelerinin düz bir şekilde sıralandığı (özne başta, yüklem sonda gibi) cümle türüne kurallı cümle denir. Yaxshilik qilishni hayotining bir parchasiga aylantirgan (İyilik yapmayı hayatının bir parçası haline getirmiş). Kurallı cümlesinin çözüm ağacı aşağıdaki gibidir.



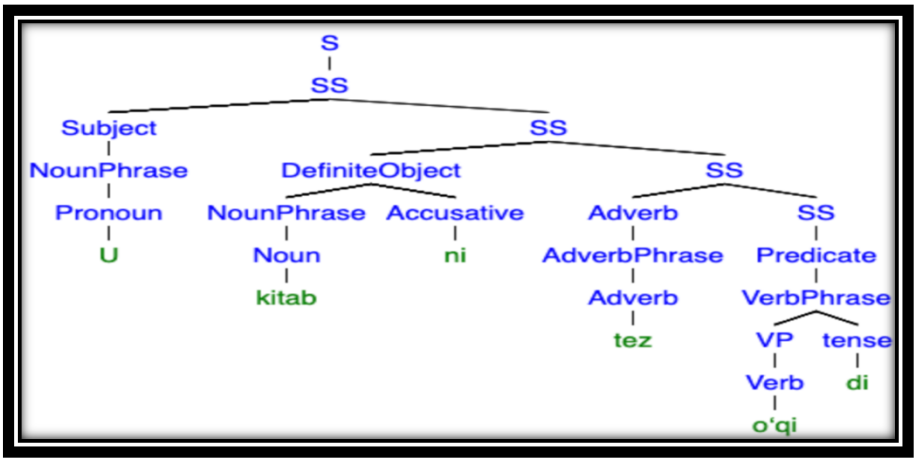
Şekil 4.21: Örnek kurallı cümlenin çözüm ağacı.

Devrik cümlede, yüklem cümlenin sonunda olmayan bir cümle türüdür. Men oldim kitobni (Ben aldım kitabı). Devrik cümlesinin çözüm ağacı aşağıdaki gibidir.



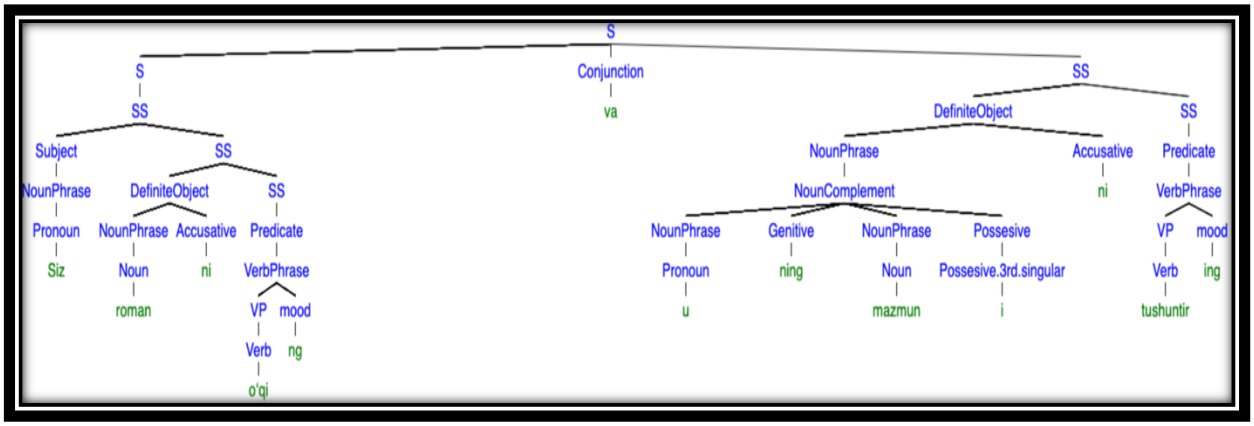
Şekil 4.22: Örnek devrik cümlenin çözüm ağacı.

Özbekçe basit cümle örneği ve çözüm ağacı aşağıda sunulmuştur. U kitobni tez o'qidi (O kitabı hızlı okudu).



Şekil 4.23: Örnek basit cümlenin çözüm ağacı.

Özbekçede birleşik cümle, bağlaçlarla oluşturulan cümlelerdir. Siz romanni o'qing va uning mazmunini tushuntiring (Romanı okuyun ve onun içeriğini anlatın).



Şekil 4.24: Örnek birleşik cümlenin çözüm ağacı.

4.3.6. Özel Durumlar, İstisnalar ve Önerilen Çözümler ile Özbekçenin Nuve Parser ile Sözdizimsel Çözümlemesi

Özbek alfabesinin tarih boyunca değişimi ve dilde bulunan kuralsız durumlar dil bilgisini, kelime ve dolayısıyla cümle yapısını etkilemiştir. Biçimbilimsel çözümleme esnasında karşılaşılan bazı istisnalar ve özel durumlar dolaylı olarak sözdizimsel çözümlemeyi de etkilemektedir. Ayrıca sözdizimsel çözümlemede eklerin ayrı ya da bitişik yazılması gibi temel durumlar direkt olarak sözdizimsel çözümlemeyi ve cümle yapısını etkilemektedir. Bu bölümde Özbekçe sözdizimsel çözümlemesi esnasında karşılaşılan özel durumlar, istisnalar ve Nuve Parser uygulama ortamında önerilen çözümler sunulmaktadır.

Özbekçede biçimbilimsel ve sözdizimsel çözümlemede çözülmesi gereken özel bir durum soru ekinin bitişik yazılmasıdır. Özbekçede soru eki ayrı yazılmaz. Örneğin "bolasimi (çocuğu mu)?" örneğinde bola (çocuk) gövdesine eklenen "mi" soru eki hem biçimbilimsel çözümlemede belirtilmeli hem de sözdizimsel çözümlemede ayrı yazılmayıp çözüm ağacında ayrıca gösterilmesi için, biçimbilimsel çözümlerde kelime ile bitişik yazılan soru eki, sözdizimsel çözüm için kullanılan gramer kurallarında ayrıca tanımlanmıştır. Ayrıca soru ekinin bitişik yazılması sözdizimsel çözümleme için biçimbilimsel çözümleme ihtiyacını da ortaya koymaktadır. Yüklem bir fiil unsurundan oluşabileceği gibi, bir fiil unsuru ile soru unsurunun birleşiminden de oluşabilir. Soru unsurlarında soru ekinden sonra şahıs eklerinin gelmesi gibi karmaşık durumlar oluşabilir. Bu gibi durumlar için biçimbilimsel çözümlemede ve gramer kurallarında soru ekinin ayrıca irdelenmesi gerekliliği ortaya çıkmıştır. Nuve Parser gramer tanımlamalarında ve biçimbilimsel çözümlemelerde soru ekinin bitişik yazılması direk ve dolaylı olarak bazı istisnai durumların kontrol edilmesi gerekliliğini ortaya koymuştur.

Özbekçede yine soru ekinin dolaylı olarak içinde bulunduğu yazılışları birebir aynı olan iki kelimenin anlamları ve kök kelime türlerinin farklı olduğu istisnalara rastlanmaktadır. Örneğin bekmi, kök kelimenin sıfat olduğu ve soru ekinin eklendiği biçimde ya da kök kelimenin isim olduğu ve soru ekinin eklendiği biçimde oluşabilmektedir. Aşağıda gösterilen örnek hem biçimbilimsel hem de sözdizimsel çözümleme için istisnai bir durumdur.

Bekmi (Bey mi)?

İsim (Bek)+3. Tekil Şahıs Eki + Soru Eki(mi)

Bekmi (Kapalı mı)?

Sıfat (Bek)+3. Tekil Şahıs Eki + Soru Eki(mi)

Yukarıdaki gibi istisnai durumların hem biçimbilimsel çözümlemede hem de kök kelime türü de değiştiği için sözdizimsel çözümlemede özel tanımlamalarının yapılması gerekmektedir.

Yine benzer şekilde Özbekçede aynı kelimelerden oluşan bazı cümleler durum ve ifadenin belirtilmesi için kullanılan noktalama işaretleri ile farklı anlamlara (ünlem ya da soru) sahip olabilir. Örneğin; “Qancha kishi (Kaç kişi)?” ve “Qancha kishi (Nice/Çok kişi)!” bu türde istisnalar özel durumlar olarak tanımlanıp cümlenin sonundaki noktalama işareti verisetine eklenerek ayırt edilmektedir.

Özbekçede bir diğer istisnai durum olumsuzluk ekinde gözlemlenmektedir. Olumsuzluk anlamı fiillerde -ma olusuzluk eki ile, isimlerde ise siz (sız), yoq (yok) ve emas (değil) olumsuzluk edatları ve eki ile yapılır. Burada sözdizimsel çözümlemede de karşımıza çıkan istisnai durum emas ve yoq eklerinin (yardımcı kelimelerinin) ayrı yazılmasıdır. Eğer tanımlaması doğru bir şekilde yapılmasa sözdizimsel çözümleyici emas ve yoq yardımcı kelimeleri ile önceki kelimenin ayrı yazılmasından dolayı ayrı iki parça olarak ele alarak çözümleme yapmaktadır. Bunun için emas ve yoq yardımcı kelimeleri istisnai bir durum olarak tanımlanmalıdır. Ayrıca emas ve yoq yardımcı kelimelerinden sonra gelen şahıs eklerinin hem biçimbilimsel hem de sözdizimsel çözümleme için ayrıca irdelenmesi ve istisnai durumla birlikte tanımlanması gerekmektedir.

İstisnai bazı durumlarda çoğunlukla edebi eserlerde belirtme hâli eki, ilgi hâli fonksiyonunu üstlenerek belirtili isim tamlaması oluşturabilmektedir. Bu istisnai durumda özel durumlara eklenerek sözdizimsel çözümleme için ayrıca tanımlanmıştır.

Türkçede sıfat fiil tamlamasında iyelik ekleri tamlayana (sıfat fiil ekine) gelirken, Özbekçede tamlanana gelmektedir. Örneğin, uyg'ongan paytlarim (uyandığım zamanlar) tamlamasında bu durum gösterilmektedir. Bu gibi özel durumlar da ayrıca Türkçeden aşina olunan yapıdan farklı olarak Özbekçe için özel bir durum olarak tanımlanmıştır.

Özbekçede bazı metinlerde yer alan ikilemelerde ikileme arasına konulan "-" işareti ayrıştırılarak iki kelime ikileme olarak tanımlanıp çözümlenmesi sağlanmıştır. Veriseti tanımlamaları çözümlemelerin doğru olması için bir modelde tanımlanmış ve Nuve Parser metin giriş alanında kullanıcıların bu türde özel işaret kullanmaları kısıtlanarak, kullanıcılar uyarılmıştır.

Verisetinde yer almayan özellikle bazı özel isimlerden dolayı sözdizimsel çözümlemenin yapılamadığı gözlemlenmiştir. Bu kelimelerin verisetinde tanımlanması sonrası sözdizimsel çözümlerin sağlandığı gözlemlenmiştir. Çalışmanın başarısı verisetinin büyüklüğü ile doğru orantılı olarak artmakta olup, özel durumların ve istisnaların çözümlemeleri ile istenilen sonuçlar alınmıştır.

5. TARTIŞMA VE SONUÇ

Bu tezde Özbekçenin biçimbilimsel çözümlemesi, biçimbilimsel belirsizlik giderme ve sözdizimsel çözümleme konuları detayları ile anlatılmıştır. Önerilen yöntemler uygulamalar ile test edilerek sonuçları ve analizleri bu bölümde paylaşılmıştır. Araştırmalar ve uygulamaların her birinin birbirleriyle ilişkili olması nedeniyle öncelikli olarak biçimbilimsel çözümlemede, sonra biçimbilimsel belirsizlik gidermede ve son olarak sözdizimsel çözümlemede sırasıyla istenilen sonuçlar alınarak bir sonraki aşamaya geçilmiştir. Bu çalışma için Özbekçe grameri detayları ile araştırılmış, düzenli kurallar, istisnalar ve özel durumlar ayrı ayrı yönetilmiştir. Birçok kaynak taranarak, Latin alfabesine geçiş sonrası hazırlanan kaynaklar çoğunluk ile kullanılmış, ihtiyaç duyulduğunda eski kaynaklardan da kontrol amaçlı faydalanılmıştır.

İlk olarak iki düzeyli Özbek morfolojisi detayları ile araştırılmıştır. İki düzeyli ortografi (yazım) kuralları tanmlanarak, Özbekçenin morfolojik (biçimbilimsel) çözümlemesi sağlanmıştır. Biçimbilimsel çözümleme esnasında karşılaşılan özel durumlar ve istisnalar araştırılmış ve uygulama aşamasında tanımlanmıştır. Çalışmanın uygulaması bu çalışma için geliştirilen Nuve Studio uygulama aracı ile yapılmıştır.

Özbekçenin biçimbilimsel çözümlemesinin uygulama aşamasında, 30.000 kök kelime ve 370 ek ile tanımlı kök ve ek sözlüğü kullanılmıştır. Uygulama için 26 iki düzeyli ortografi kuralı tanımlanarak bu kurallar doğrultusunda çözümlemeler yapılmıştır. İstisna durumlar ayrıca işlenerek sözlüğe eklenmiş ve çözümlemeleri sağlanmıştır. Bir kelimenin birden fazla biçimbilimsel çözümünün olması biçimbilimsel belirsizlik gidermenin konusu olarak biçimbilimsel belirsizlik giderme aşamasında sonuçları yorumlanmıştır. Bir morfolojik çözümleyici olarak ürettiği sonuçların doğruluğu karşılaştırılmıştır. Bu bağlamda Nuve Studio sözlükte tanımlı olan kelimelerin biçimbilimsel çözümlemesi için tekniğin bilinen durumuna göre kabul edilebilir oranda başarılı sonuçlar ortaya koymuştur. Bir kelimenin birçok çözümlemesinin olduğu durumlarda üretilen çözümlerin doğru olduğu gözlemlenmiştir. Uygulama testlerinde tüm sözlük kullanılarak sağlanan testte 30.000 kelimenin 27.993 tanesi doğru (true) ve 2.007 tanesi yanlış(false) çözümlenerek %93,31 oranında bir başarı ile istenilen sonuca ulaşmıştır.

Özbekçenin biçimbilimsel belirsizlik giderme ihtiyacı biçimbilimsel çözümleme sonrasında ortaya çıkmıştır. Birçok kelimenin birden fazla çözümünün olması ve verilen bir cümle içerisinde en doğru çözümün bulunması biçimbilimsel belirsizlik gidermenin konusudur.

Nuve Studio çıktıları MorphNet modelinin girdi formatına uygun olarak düzenlenerek ilk aşamada 13.000 kelimelik bir test yapılmıştır. Model yapısı gereği otomatik olarak veri setini üç parçaya bölerek bunun %80'i eğitim, %10'u geliştirme (development) ve %10'u test için kullanılmaktadır. Buna göre veri setinde bulunan kelimelerin 10.400 tanesi eğitim, 1300 tanesi geliştirme ve 1300 tanesi test için kullanılmaktadır. Modelin başarısının harici olarak ölçümlenmesi için 300 kelime harici tutularak özellikle veri setinden seçilmeden testleri yapılmış ve sonuçları paylaşılmıştır. Veri setinde yer alan her bir kelimenin biçimbilimsel çözümlemeleri içerisinde en doğru olan çözümler elle seçilerek ilk çözüme yazılmış ya da bir tane biçimbilimsel çözümü bulunan kelimeler için en doğru çözüm kabul edilmiştir. MorphNet modeline uygun olarak girdiler tamamlandıktan sonra eğitime ve teste başlanmıştır. Testler, i7 işlemcili, Nvidia ekran kartlı 8 GB GPU, 16 GB RAM ve 256 GB SSD içeren bir bilgisayar ile 13.000 kelime içeren bir veri seti için 8-12 saat arasında iki farklı eğitim (train) yapıldıktan sonra yapılmıştır. Bu testler neticesinde 13.000 kelimeden 9.919 kelime en doğru biçimbilimsel çözüm ile çıktı vermiştir. Kalan 3.081 kelime yanlış, rastgele veya boş olarak çözümlendiği için yanlış olarak kabul edilmiştir. Daha fazla eğitim ve veri seti boyutunun artırılması ile bu başarı oranının artırılabileceği ön görülmüş olup bu test neticesinde %76,3 oranında başarı sağlanmıştır. Veri setinde yer almayan 300 kelime ayrıca bulundukları cümleler ile test edilmiş ve 300 kelimeden 168 tanesi için istenilen sonuçlar elde edilmiş, 132 tanesi ilk biçimbilimsel sonucu ile eşleşmemiş, rastgele çözümlenmiş ya da boş olarak çözümlenerek %56 oranında doğruluk sağlanmıştır.

Veri seti boyutunun artırılması ile başarı oranının artacağı ön görüsüyle, sonraki test için 30.000 kelimeden oluşan veri seti kullanılmıştır. Bu testte de ilk testte kullanılan bilgisayar kullanılmış olup, eğitim süresi veri seti boyutu ile doğru orantılı olarak biraz daha fazla sürede tamamlanmıştır. Belirli eğitim süresi sonrasında 30.000 kelimenin 24.248 tanesi doğru olarak çözümlenmiş, 5.752 tane çözüm yanlış, rastgele ya da boş olarak çözümlenmiştir. Bu testler sonucunda %80,83 oranında başarı sağlanmıştır. Ortaya çıkan bu sonuç veri seti boyutunun artırıldıkça ve yeterli eğitim sağlandıkça başarı oranının artacağı ön görüsünü doğrulamıştır.

Özbekçenin biçimbilimsel çözümlemesi ve biçimbilimsel belirsizlik gidermesi sonrasında elde edilen veriler sözdizimsel çözümlemede kullanılacak formata çevrilerek, Özbekçenin sözdizimsel çözümlemesi üzerine çalışmalar yapılmıştır. Özbekçenin sözdizimsel çözümlemesi yapılarak, sözdizimsel çözümleme bu tez kapsamında geliştirilen Nuve Parser ile gerçekleştirilip test edilmiştir. Özbekçenin sözdizimsel çözümlemesi için Özbekçe grameri bağlamdan bağımsız gramer(context free grammar) kuralları ile ifade edilmiştir. Özbekçe grameri 286 bağlamdan bağımsız gramer kuralı ile ifade edilerek, sözdizimin çözüm ağaçlarının Özbekçe gramerine uygunluğunu ve geçerliliğini kontrol eden 24 doğrulama kuralı ayrıca yazılmıştır.

Biçimbilimsel çözümleme aşamasında olduğu gibi sözdizimsel çözümleme aşamasında da özel durumlar ve istisnalar ayrıca ele alınmıştır. Özel durumların ve istisnaların yönetimi önceki bölümde detaylı olarak anlatılmış olup, test sonuçlarında kullanılan veri içerisinde özel durumlar ve istisnalar yer almaktadır. 30.000 kelimelik veri seti kullanılarak testler yapılmış ve Özbekçe sözdizimi açısından her bir cümle türü ve kelime grubu test edilerek istenilen sonuçlar alınmıştır. Bir cümle için birden fazla sözdizimsel çözüm üretilmesi sözdizimsel belirsizlik konusu olarak değerlendirilmiş olup, doğrulama kurallarına göre geçerli olmayan çıktılar uygulama aşamasında çıktıya yansıtılmamıştır.

Bu tezde, kural tabanlı Özbekçe biçimbilimsel ve sözdizimsel çözümlemesi yapılıştır. Önerilen yöntemlerin uygulama ve testleri neticesinde istenilen düzeyde ve tekniğin bilinen durumuna göre kabul edilebilir oranda başarılı sonuçlar elde edilmiştir.

Nuve Studio ve Nuve Parser uygulamalarının geliştirilmesinin ve birleştirilmesinin yanında biçimbilimsel belirsizlik giderme için kullanılan uygulamanın birleştirilmesi ve hepsinin online bir platformda başta Türki diller olmak üzere birçok dilde bu konuda çalışan araştırmacıların faydalanması için kullanıma açılması hedeflenmiştir. Bu doğrultuda Nuve Studio[4] online bir versiyon olarak kullanıma açılmış ve açık kaynak kodlu (KÜÇÜK, 2021) olarak paylaşılmıştır.

[4] www.nuvestudio.com

KAYNAKLAR

ABDURAKHMONOVA, N.Z., 2019, A Two-Level Morphological Analysis of the Uzbek Corpus, *Иностранная филология, Социальная и национальная вариативность языка и литературы*, 425-430.

Аскарова, Р.Н., 1972, О котовниках Кавказа, *Материалы по флоре и систематике высших растений Азербайджана*, Bakü, 72-110.

ADALI, E., 2016, Doğal Dil İşleme, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5 (2). https://dergipark.org.tr/en/pub/tbbmd/issue/22245/238797, [Ziyaret tarihi: 5 Aralık 2021].

AGIRRE, E., ALEGRIA, I., ARREGI, X., ARTOLA, X., DIAZ de ILLARAZA, A., MARITXALAR, M., SARASOLA, K., URKIA, M., 1993, XUXEN: A spelling checker/corrector for Basque based on Two-Level morphology, *Third Conference on Applied Natural Language Processing*, 119-125.

AKYÜREK, E., 2019, *Morse*, https://github.com/ekinakyurek/Morse.jl, [Ziyaret tarihi: 10 Aralık 2021].

ALAM, Y.S., 1983, A Two-Level Morphological Analysis of Japanese, *In: Texas Linguistic Forum,* 22, 229-252.

ANTWORTH, E.L., 1990, PC-KIMMO: A Two-level Processor of Morphological Analysis, *Summer Institute of Lingustics*, Dallas, TX.

ATADJANOV, J.A., 2016, Models of Morphological Analysis of Uzbek Words, *Кибернетика и программирование*, 6, 70-73.

DAYANIK, E., AKYÜREK, E. and YURET, D., 2018, Morphnet: A sequence-to-sequence model that combines morphological analysis and disambiguation, *arXiv preprint arXiv:1805.07946*. https://arxiv.org/abs/1805.07946, [Ziyaret tarihi: 5 Aralık 2021].

ERCİLASUN, A., 2007, *Türk Lehçeleri Grameri,* Akçağ Yayınları, Ankara, ISBN: 978-975-338-885-6.

HURSKAINEN, A., 1992, A Two-Level Computer Formalism for the Analysis of Bantu Morphology An Application to Swahili, *In: NJAS*, 1 (1), 87-119.

İMER, K., KOCAMAN, A., ÖZSOY, S., 2011, *Dilbilim Sözlüğü*, Boğaziçi Üniversitesi Yayınevi, İstanbul, ISBN: 978-605-4238-53-8.

KARAHAN, L., 2012, *Türkçede Söz Dizimi*, Akçağ Yayınları, Ankara, ISBN: 978-975-338-279-3.

KIM, D.B., LEE, S.J., CHOI, K.S., KIM, G.C., 1994, A two-level morphological analysis of Korean, *COLING The 15th International Conference on Computational Linguistics,* 1(1), 535-539.

KOSKENNIEMI, K., 1983, Two-level model for morphological analysis, *In: IJCAL*, 683-685.

KOSKENNIEMI, K., 1984, A general computational model for word-form recognition and production, *Proceedings of the 4th Nordic Conference of Computational Linguistics (NODALIDA 1983)*, 145-154.

KÜÇÜK, A.E., 2021, *Nuve Studio Source Code*, https://github.com/kucukea/nuve-studio, [Ziyaret tarihi: 5 Aralık 2021].

LEWIS, D. and JONES, K., 1996, Natural Language Processing for Information Retrieval, *Communications of the ACM*, 39 (1), 92-101.

MATLATIPOV, G. and VETULANI, Z., 2009, Representation of Uzbek Morphology in Prolog, *In: Aspects of Natural Language Processing*, Springer, Berlin, Heidelberg, 83-110.

NURMANOV, A., MAXMUDOV, N., AXHEDOV, A., SOLIHZAEVA, S., 1992, *O'zbek Tilinin Mazmuniy Sintaksisi*, Özbekistan Cumhuriyeti Bilimler Akademisi Yayınevi, Taşkent, Özbekistan.

OFLAZER, K., 1994, Two-level description of Turkish morphology, *Literary and Linguistic Computing,* 9(2), 137-148.

SHANG, M., 2011, Syntax Tree Generator, http://mshang.ca/2011/10/28/syntax-tree-generator.html, [Ziyaret tarihi: 10 Aralık 2021].

SIPSER, M., 1997, Context-free grammars, *Introduction to the Theory of Computation*, 91-122.

SPROAT, R., 1992, *Morphology and Computation*, Mit Press, Cambridge, ISBN: 978-0262193146.

SULAYMONOV, A., XOLİYOROV, X., OMONTURDİEV, J., 1979, *Hozirgi O'zbek Adabiy Tili. Sintaksis*, Toşkent: O'qituvchi, Taşkent, Özbekistan.

ŞENTÜRK, F. ve ADALI, E., 2016, Yetkin Bul ve Değiştir, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 3(1), 63-71. https://dergipark.org.tr/en/pub/tbbmd/issue/22242/238773, [Ziyaret tarihi: 5 Aralık 2021].

ŞİMŞEK, R., 1987, *Örneklerle Türkçe Sözdizimi (Turkish Syntax with Examples)*, Kuzey Matbaacılık, Trabzon.

THE UZBEKISTAN CABINET of MINISTER'S RESOLUTION, 1995, Principal Orthographic Rules For The Uzbek Language, 24 August 1995 Tashkent, Uzbekistan, No. 339.

TRIVEDI, A., PANT, N., SHAH, P., SONIK, S., 2018, Speech to text and text to speech recognition systems-Areview, *IOSR J. Comput. Eng,* 20 (2), 36-43.

TOLEU, A., TOLEGEN, G., MAKAZHANOV, A., 2017, Character-aware neural morphological disambiguation. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Volume 2, 666-671.

TÜLEK, M., 2007, *Türkçe için Metin Özetleme,* Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

ÜŞENMEZ, E., 2012, *Modern Özbek Türkçesi: (Giriş-Özbekistan-Dil Tarihi-Metinler-Gramer-Sözlük)*, Akademik Kitaplar, İstanbul, Türkiye, ISBN: 978-605-568-850-9.

ÜŞENMEZ, E., 2017, Özbek Türkçesinin Oluşumu, *Turkish Studies (Elektronik),*12 (7), 463-476.

ШОАБДУРАҲМОНОВ, Ш., АСҚАРОВА, М., ҲОЖИЕВ, А.Р.И., ДОНИЁРОВ, Х., 1980, Ҳозирги ўзбек адабий тили, Ўқитувчи, 114.

YAMAN, E., 1994, *Türkiye Türkçesi ve Özbek Türkçesi'nin söz dizimi bakımından karşılaştırılması*, Doktora Tezi, Gazi Üniversitesi Sosyal Bilimler Enstitüsü.

YAMAN, E., 2000, *Türkiye Türkçesi ve Özbek Türkçesi'nin söz dizimi bakımından karşılaştırılması*, Atatürk Kültür, Dil ve Tarih Yüksek Kurumu Türk Dil Kurumu Yayınları:742, Ankara, ISBN: 978-975-167-420-3.

YİNER, Z., KURT, A., KULAMSHAEV, K., ZAFER, H.R., 2016, Kyrgyz Orthography and Morphotactics with Implementation in NUVE, *International Conference on Engineering and Natural Science (ICENS)*, 24-28 May 2016 Sarajevo, 1650-1658.

YUSUPOVA, N., 2018, *Türkçe Özbekçe Sözlük*, Türk Dil Kurumu Yayınları, Ankara, Türkiye.

ZAFER, H.R., 2011, *A generic syntactic parser for Turkic languages,* Master of Science, Fatih University The Graduate Institute of Sciences and Engineering.

ZAFER, H.R., 2018, *Nuve: A Natural Language Processing Library for Turkish in C#,* https://github.com/hrzafer/nuve, [Ziyaret tarihi: 5 Aralık 2021].

EKLER