



**T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**



DOKTORA TEZİ

KAZAKÇA’NIN BİÇİMBİLİMSEL ve SÖZDİZİMSEL ANALİZİ

Züleyha YİNER

**DANIŞMAN
Doç. Dr. Atakan KURT**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

İSTANBUL-2021

Bu çalışma, 25.10.2021 tarihinde aşğıdaki jüri tarafından Bilgisayar Mühendisliğı Anabilim Dalı, Bilgisayar Mühendisliğı Programında 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

Doç. Dr. Zeynep ORMAN
İstanbul Üniversitesi-Cerrahpaşa
Mühendislik Fakültesi

Doç. Dr. Yılmaz KAYA
Siirt Üniversitesi
Mühendislik Fakültesi

Doç. Dr. Taner ÇEVİK
Nişantaşı Üniversite
Mühendislik Mimarlık 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

Doktora eğitimimin her aşamasında bilgi ve tecrübelerini esirgemeyen, her zaman yol gösteren çok değerli hocam ve tez danışmanım Sayın Doç. Dr. Atakan KURT'a teşekkürü bir borç bilirim.

Tez izleme süreçlerinde vermiş oldukları destek ve önerilerinden ötürü çok değerli tez izleme komitesi üye hocalarım Sayın Doç. Dr. Zeynep ORMAN ve Sayın Doç. Dr. Yılmaz KAYA hocalarıma sonsuz teşekkürlerimi sunarım.

Başta öğrenim hayatım olmak üzere tüm hayatım boyunca her türlü maddi ve manevi desteğini aldığım başta annem ve babam olmak üzere tüm kardeşlerime sonsuz teşekkürlerimi sunarım.

Gerek doktora sürecinde gerek sosyal hayatımda yanımda olan ve adını burada yazamadığım tüm arkadaşlarıma sonsuz teşekkür ederim.

Ekim 2021

Züleyha YİNER

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ.....	ix
SİMGE VE KISALTMA LİSTESİ	x
ÖZET	xi
SUMMARY	xii
1. GİRİŞ	1
2. GENEL KISIMLAR.....	3
2.1. KAZAKÇA ALFABE.....	3
2.2. MORFOLOJİK ANALİZ.....	3
2.3. KAZAKÇA EK ÇEŞİTLERİ	4
2.3.1. İsim Çekim Ekleri (NNI_ xxx)	5
2.3.2. İsim Yapım Ekleri (NxD_ xxx)	7
2.3.2.1. İsimden İsim Yapan Yapım Ekleri (NND_ xxx)	7
2.3.2.2. Fiilden İsim Yapan Yapım Ekleri (NVD_ xxx)	7
2.3.2.3. Sıfattan İsim Yapan (NJD_ xxx) ya da Sıfattan Sıfat Yapan Yapım Ekleri (JJD_ xxx)	8
2.3.2.4. Fiilden Sıfat Yapan Yapım Eki (JVD_ xxx)	8
2.3.3. Fiil Çekim Ekleri (VVI_ xxx)	9
2.3.4. Fiil Yapım Ekleri (VxD_ xxx)	12
2.3.4.1. İsimden Fiil Yapan Yapım Eki (VND_ xxx)	12
2.3.4.2. Fiilden Fiil Yapan Yapım Eki (VVD_ xxx)	12
2.3.4.3. Sıfattan Zarf Yapan Yapım Eki (AJD_ xxx)	13
2.3.4.4. Fiilden Zarf Yapan Yapım ekleri (AVD_ xxx)	13
3. MALZEME VE YÖNTEM.....	14
3.1. KAZAKÇA BİÇİMBİLİMSEL ANALİZ	14
3.1.1. Biçimbilimsel Çözümleme	14
3.1.2. İki Düzeyli Morfoloji	16
3.1.3. Kazakçanın Biçimbilimsel Çözümlemesi.....	17

3.1.3.1. İki Düzeyli Yazım Kuralları	17
3.1.3.2. Morfotaktik	26
3.2. MORFOLOJİK BELİRSİZLİK GİDERME	32
3.3. KAZAKÇA SÖZDİZİMSEL ANALİZ	35
3.3.1. Context Free Gramer (Bağlamdan Bağımsız Gramer)	36
3.3.2. Kazakçanın Gramer Yapısı Bakımından İncelenmesi.....	37
3.3.3. Kazakça Kelime Grupları ve Sözdizimsel Analizleri.....	38
3.3.3.1. İsim Tamlaması (Noun complement)	38
3.3.3.2. Sıfat Tamlaması (Adjective Complement)	39
3.3.3.3. Sıfat-Fiil Grubu (Verbal Adjective Group).....	40
3.3.3.4. İsim-Fiil Grubu (Verbal Noun Group)	41
3.3.3.5. Zarf-Fiil Grubu (Verbal Adverb Group)	41
3.3.3.6. Tekrar Grubu (Reduplication Group)	42
3.3.3.7. Edat Grubu (Postposition Group)	43
3.3.3.8. Bağlama Grubu (Conjunction Group).....	44
3.3.3.9. Unvan Grubu (Title Group).....	45
3.3.3.10. Özel İsim Grubu (Person Name Group).....	45
3.3.3.11. Ünlem Grubu (Interjection Group)	45
3.3.3.12. Sayı Grubu (Number Group)	46
3.3.3.13. Birleşik Fiil (Compound Verb Group).....	46
4. BULGULAR.....	48
4.1. KAZAKÇA BİÇİMBİLİMSEL ANALİZİNİN NÜVE'DE GERÇEKLENMESİ.....	48
4.1.1. Nüve Biçimbilimsel Çözümleyici	48
4.1.2. Nüve'de Kazakça Morfolojik Analiz Örnekleri	53
4.1.3. Kazakça Biçimbilimsel Analiz Sırasında Bulunan Özel Durumlar ve İstisnalar	55
4.2. KAZAKÇA SÖZDİZİMSEL ANALİZİNİN NÜVE'DE GERÇEKLENMESİ.....	56
4.2.1. Nüve Sözdizimsel Çözümleyici	57
4.2.2. Nüve'de Kazakça Sözdizimsel Analiz Örnekleri	64
4.2.3. Kazakça Sözdizimsel Analizi Sırasında Bulunan Özel Durumlar ve İstisnalar	70
5. TARTIŞMA VE SONUÇ	71
KAYNAKLAR.....	74
ÖZGEÇMİŞ	78

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Morfolojik ve sözdizimsel analiz akış diyagramı.	4
Şekil 3.1: İsim gövdeleri için SDM modeli (kapsamlı).	28
Şekil 3.2: İsim gövdeleri için SDM modeli (özet).	29
Şekil 3.3: Fiil gövdeleri için SDM modeli birinci kısım.	31
Şekil 3.4: Fiil gövdeleri için SDM modeli ikinci kısım.	32
Şekil 3.5: "Elif güzel yazı yazar" cümlesi için çözüm (parse) ağacı.	37
Şekil 3.6: Belirtisiz isim tamlaması.	38
Şekil 3.7: Belirtili isim tamlaması.	39
Şekil 3.8: Sıfat tamlaması.	40
Şekil 3.9: Sıfat-fiil grubu.	40
Şekil 3.10: İsim-fiil grubu.	41
Şekil 3.11: Zarf-fiil grubu.	42
Şekil 3.12: Tekrar grubu.	43
Şekil 3.13: Edat grubu.	44
Şekil 3.14: Bağlama grubu.	44
Şekil 3.15: Unvan grubu.	45
Şekil 3.16: Özel isim grubu.	45
Şekil 3.17: Ünlem grubu.	46
Şekil 3.18: Sayı grubu.	46
Şekil 3.19: Birleşik fiil grubu.	47
Şekil 4.1: Ortografik kurallar XML dosyası.	51
Şekil 4.2: Morfotaktik XML dosyası.	52
Şekil 4.3: Morfolojik çözüm örneği.	54

Şekil 4.4: Sözdizimsel çözümleyici akış diyagramı.	57
Şekil 4.5: Şekil 4.3’de verilen morfolojik çözümler örneğinin ek ve tag eşleşmesinden sonraki hali.	60
Şekil 4.6: Gramer.XML (Özne).....	61
Şekil 4.7: Gramer.XML (Nesne).	61
Şekil 4.8: Gramer.XML (Yüklem).	62
Şekil 4.9: Gramer.XML den bir bölüm (İsim unsuru-NP).	63
Şekil 4.10: Validation.XML dosyasından bir kesit.	64
Şekil 4.11: Olumlu isim cümlesi.	65
Şekil 4.12: Olumsuz isim cümlesi.	65
Şekil 4.13: Soru isim cümlesi.	66
Şekil 4.14: Olumlu fiil cümlesi.	66
Şekil 4.15: Olumsuz fiil cümlesi.	67
Şekil 4.16: Soru fiil cümlesi.	67
Şekil 4.17: Kurallı cümle.....	68
Şekil 4.18: Devrik cümle.	68
Şekil 4.19: Basit cümle.....	69
Şekil 4.20: Birleşik cümle.	69

TABLO LİSTESİ

Sayfa No

Tablo 2.1: Kazak alfabesi.	3
Tablo 2.2: Kazakça ünlüler.	3
Tablo 2.3: İsim çekim ekleri.	6
Tablo 2.4: İsimden isim yapan yapım ekleri.	7
Tablo 2.5: Fiilden isim yapan yapım ekleri.	8
Tablo 2.6: Sıfattan isim yapan ya da sıfattan sıfat yapan yapım ekleri.	8
Tablo 2.7: Fiilden sıfat yapan yapım eki.	9
Tablo 2.8: Fiil çekim ekleri.	9
Tablo 2.9: İsimden fiil yapan yapım ekleri.	12
Tablo 2.10: Fiilden fiil yapan yapım eki.	12
Tablo 2.11: Sıfattan zarf yapan yapım eki.	13
Tablo 2.12: Fiilden zarf yapan yapım ekleri.	13
Tablo 3.1: İsim gövdelerine gelen çekim ekleri	27
Tablo 4.1: Ekler sözlüğü.	49
Tablo 4.2: Kökler sözlüğü.	49
Tablo 4.3: Manuel tag ve eklerin eşleştirilmesi.	58

SİMGE VE KISALTMA LİSTESİ

Kısaltmalar	Açıklama
CG	: Conjunction Group
CVG	: Compound Verbs
DObj	: Direct Object
HMM	: Hidden Markov Model
IG	: Interjection Group
IObj	: Indirect Object
JC	: Adjective complement
LC	: Left Context
NC	: Noun Complement
NG	: Number Group
NP	: Noun Phrase
PG	: Postposition Group
PNG	: Person Name Group
POS	: Part of Speech Tagging
Pre	: Predicate (Yüklem)
RC	: Right Context
RG	: Reduplication Group
SDM	: Sonlu Durum Makinası
VAG	: Verbal Adverb Group
VJG	: Verbal Adjective Group
VNG	: Verbal Noun Group

ÖZET

DOKTORA TEZİ

KAZAKÇANIN BİÇİMBİLİMSEL ve SÖZDİZİMSEL ANALİZİ

Züleyha YİNER

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman : Doç. Dr. Atakan KURT

Bu tezde Türki dillerden biri olan Kazakçanın biçimbilimsel (morfolojik) ve sözdizimsel analizi yapılmıştır. Tezin ilk aşaması olan biçimbilimsel analizde iki düzeyli morfoloji modeli kullanılmıştır. Bu modelde yazım kuralları iki düzeyli kurallar şeklinde yazılır ve dilin morfolojik kuralları (dilde bulunan eklerin sıralamasını gösteren kurallar) Sonlu Durum Makinaları olarak kodlanır. Kazakçanın iki düzeyli morfolojisinin test edilmesi için dilden bağımsız ve açık kaynaklı morfolojik çözümleyici ve makine çevirisi sistemi olan Nüve kullanılmıştır. Kazakçanın biçimbilimsel çözümlemesi (morfolojik üretim ve analizi) web ortamına taşınarak kullanıcılara açık hale getirilmiştir (<https://nuvestudio.com/>). Tez çalışmasının ikinci adımında Kazakçanın sözdizimsel analizi yapılarak Kazakçanın grameri CFG (Context Free Grammar) kuralları ile ifade edilmiştir. Bu analiz sırasında Kazakçanın tüm cümle türleri ve kelime grupları (isim tamlaması, sıfat tamlaması, sıfat-fiil, isim-fiil, zarf-fiil vb.) ayrı ayrı çalışılmıştır. Daha sonra Kazakça gramer kuralları Nüve sözdizimsel çözümleyicisinde gerçekleştirilmiş ve test edilmiştir. Nüve sözdizimsel çözümleyicisi, verilen Kazakça bir cümleyi Kazakça gramer kurallarına göre öğelerine ayrıştırarak olası tüm çözümleri üretebilmekte ve üretilen çözümleri bir ağaç olarak görüntüleyebilmektedir. Kazakça gramer kurallarının testi için yaklaşık 1100 cümleden oluşan bir veri hazırlanmış ve veri setini bir bütün olarak test edebilme özelliği Nüve'ye bir modül olarak eklenmiştir.

Ekim 2021, 91 sayfa.

Anahtar kelimeler: Doğal dil işleme, Morfolojik Analiz, Sözdizimsel Analiz

SUMMARY

Ph.D. THESIS

MORPHOLOGICAL AND SYNTACTIC ANALYSIS OF KAZAKH

Zuleyha YINER

Istanbul University-Cerrahpasa

Institute of Graduate Studies

Department of Computer Engineering

Supervisor : Assoc. Prof. Dr. Atakan KURT

In this thesis, the morphological and syntactic analysis of Kazakh, a Turkic languages, is studied. In the morphological analysis, which is the first part of the thesis, a two-level morphology model is used. In this model, orthographic rules are written as two-level rules and morphotactics rules (the rules that show the order of the suffixes) are encoded as Finite State Machines. Nuve, a language-independent and open-source morphological analyzer and machine translation system, is used to implement two-level morphology of Kazakh. The morphological generation and parsing of Kazakh language is made available on the web (<https://nuvestudio.com/>). In the second part of the thesis, the syntactic analysis of Kazakh language is studied and the grammar of Kazakh is expressed with CFG (Context Free Grammar) rules. In this analysis, all sentence types and word groups (noun complement, adjective complement, verbal adjective, verbal noun, verbal adverb etc.) of Kazakh are studied. Later Kazakh grammar rules are implemented and tested on Nuve syntactic analyzer. Nuve can generate all possible solutions of a given Kazakh sentence into its syntactic parts according to Kazakh grammar rules and display the solutions as trees. A dataset consisting of nearly 1100 sentences was prepared for testing of Kazakh grammar rules and a testing module was added to Nuve to testing dataset as a whole.

October 2021, 91 pages.

Keywords: Natural Language Processing, Morphological Analysis, Syntactic Analysis

1. GİRİŞ

Dilbilimi (Linguistics) bir doğal dilin dilbilgisi yapısı, anlamını ve fonetiğini çalışan bilim koludur. Doğal dil işleme (Natural Language Processing), bilgisayar biliminin çalışma alanı olup yapay zekânın alt kollarından biridir. Doğal dil işleme (DDİ) çalışmalarında temel amaç, doğal dillerin bilgisayar ortamında tıpkı insanların anladıkları şekilde işleyebilmektir. DDİ, hesaplmalı dilbilim ya da bilgisayarlı dilbilimle (Computational Linguistics) kural tabanlı yaklaşımlar, istatistiksel tabanlı yaklaşımlar, makine öğrenmesi ya da derin öğrenme modellerini birbirine bağlar. Bu teknolojiler, bir doğal dilin metin ya da ses verisi olarak bilgisayar ortamında işlenmesini sağlamaktadır.

Doğal dil işlemede yapılan çalışmalar kelime (morphological-lexical), sözdizimsel (syntactic), anlamsal (semantic) ve söylev (pragmatic-discourse) seviyelerindedir (Şeker, 2015). DDİ uygulama alanlarına metin sınıflandırma, morfolojik analiz, kelime türü etiketleme, sözdizimsel analiz, anlamsal analiz, diller arası çeviri, karakter tanıma, yazım denetimi, soru cevaplama, ses tanıma ve bunun gibi birçok uygulama örnek olarak verilebilir.

DDİ çalışmalarında veri metin ya da sestir. Metin madenciliği sayesinde veri bilimiyle DDİ çalışmaları arasındaki ilişki sağlanır. Bir dilde verilen metine bakılarak o dilin tanınması (Kaya ve diğ., 2015, Kaya ve Ertuğrul, 2016) veya duygu analizinin yapılması (Kaya ve Özdemir, 2019), metin içerisindeki yazım yanlışlarının bulunması (Delibas, 2008) temel uygulamalara örnek olarak verilebilir. DDİ çalışmalarında ses verisinin işlenmesi, metin verisinin işlenmesine nazaran daha zordur. En yaygın ses işlenmesi çalışmalarına, metinden sese (text to speech processing) ya da tam tersi verilebilir.

Bir doğal dilin işlenmesinde elle hazırlanan kurallar dili her açıdan ele alınması noktasında sıkıntıları da beraberinde getirmektedir. Çünkü her istisnai ya da özel durumlar için kural yazılması gerekmektedir. Karşılaşılan her durum için yazılan kurallar, sistemi daha karmaşık hale getirir. Bu, hem iş yükü hem de zaman verimliliği açısından da etkili değildir. Son yıllarda yapılan çalışmalara bakıldığında yapay zekâ tabanlı modellere yoğunlaşıldığı görülmektedir. Ayrıca teknoloji çağımızda bilginin çokluğuyla modellerin de donanımsal olarak bunu karşılayabilmesi ihtiyacını da doğurmuştur. Bu yüzden yapay sinir ağları, derin öğrenme yöntemleri yaygın olarak kullanılmaya başlanmıştır. DDİ çalışmalarındaki başarının yapay

sinir ağıları modeli yaklaşımların kullanılmasıyla arttığı görülmüştür. Genelde bu çalışmalar feedforward networks (ileri beslemeli sinir ağıları), convolutional networks (evrimsel sinir ağıları), recurrent networks (tekrarlayan sinir ağıları) ve recursive networks (özyinelemeli sinir ağıları) gibi yapay sinir ağıları temellidir (Goldberg, 2016).

Türkçe için yapılmış çalışmalar yazım düzeltme, işaret kümesi tanımı, istatistiksel dil modelleme, sözdizimi ve istatistiksel çeviri olarak ana başlıkları altında verilmiştir (Oflazer, 2016).

Dil çalışmalarında önemli bir konu da bir metin içerisinde geçen kelimenin türünün belirlenmesidir. Babüroğlu ve diğ. (2019) derin öğrenme yaklaşımında Recurrent Neural Network (RNN) yapısını kullanarak Türkçe kelime türü etiketlemeyi önermiştir. Yine benzer bir yapıyla Maimaiti ve diğ. (2017) Uygurca kelime türü etiketleme çalışması yapmışlar.

DDİ çalışma alanlarından bir diğeryse makine çevirisidir. İki dil arasında yapılan çeviri sistemlerinde iki dilin de morfolojik yapısı bakımından benzerliği çalışmanın başarısını etkilemektedir. Türki dilleri arasında yapılan makine çevirisi çalışmaları olarak Kazakça-Tatarca arasında (Salimzyanov ve diğ., 2013), Kırım Tatarcası-Türkçe (ya da Türkçe-Kırım Tatarcası) arasında (Altıntaş, 2001, Gökirmak ve diğ., 2019, Sanli, 2018), Kırgızca-Türkçe (Yiner, 2016), Türkçe-Azerice (Fatullayev ve diğ., 2008, Hamzaoglu, 1993), Tatarca-Türkçe (Gilmullin, 2008), Türkçe-Türkmence (Tantuğ ve Adalı, 2018) dilleri arasında yapılan çalışmalar örnek olarak verilebilir. Kazakçadan İngilizceye (Tukeyev ve diğ., 2011, Sundetova ve diğ., 2014) çalışması da Kazakça ile diğer diller arasında makine çevirisi çalışmalarına örnek olarak verilebilir.

Kazakça için yapılan DDİ çalışmaları son yıllarda biraz daha yaygın olmaya başlamıştır. Kazakça kelimeleri çekim eklerine ayırma çalışması (Altenbek ve Wang, 2010), yazım hatası düzeltme (Makazhanov ve diğ., 2014, Kozhirbayev ve Yessenbayev, 2020), Kazakça isimlerde kök ve eklerine ayırma (morfolojik analiz) çalışması (Kairakbay ve Zaurbekov, 2013), Kazakça metinlerde normalizasyonla metin içerisindeki kelime hataların giderilmesi (Abdussaitova ve Amangeldiyeva, 2019) bu çalışmalara örnek olarak verilebilir.

2. GENEL KISIMLAR

2.1. KAZAKÇA ALFABE

Son yıllarda Kazak alfabesinin değişimi çalışmaları yaygın olarak görülmektedir. Bu tezde Kazakçanın referans alınan alfabesi Latin-Kiril karşılıkları halinde aşağıdaki tabloda verilmiştir.

Tablo 2.1: Kazak alfabesi.

Kiril	Latin	Kiril	Latin	Kiril	Latin	Kiril	Latin
Аа	a	Ии	i (y)	Рр	r	Ҫҫ	ç
Әә	ä	Йй	y	Сс	s	Шш	ş
Бб	b	Кк	k	Тт	t	Щщ	şç
Вв	v	Ққ	q	Уу	(u) w	Ъ	kesme
Гг	g	Лл	l	Ұұ	u	Ыы	ı
Ғғ	ğ	Мм	m	Үү	ü	Іі	i
Дд	d	Нн	n	Фф	f	Ь	yumuşama
Ее	e	Ңң	ñ	Хх	x	Ээ	é
Ёё	yo	Оо	o	Һһ	h	Юю	yu
Жж	j / c	Өө	ö	Цц	ts	Яя	ya
Зз	z	Пп	p				

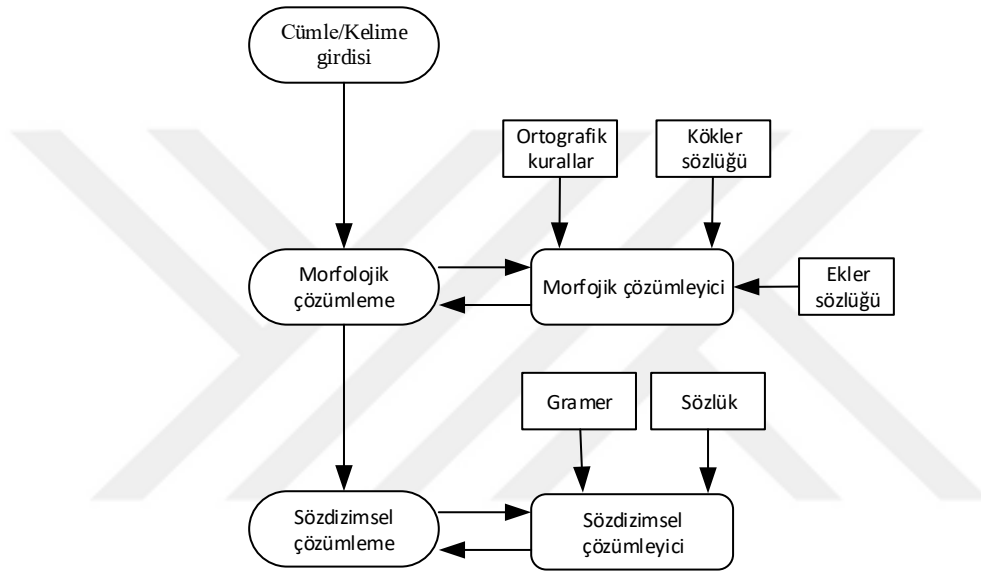
Tablo 2.2: Kazakça ünlüler.

	Düz		Yuvarlak	
İnce	e, э (e) , ә (ä)	і (i)	ө (ö)	ү (ü)
Kalın	a (a)	ы (ı)	о (o)	ұ (u)

2.2. MORFOLOJİK ANALİZ

Morfoloji kelimelerin yapısını gösterirken, morfem ise bir dilin anlamlı en küçük birimidir. Morfolojik analiz (biçimbilimsel çözümleme) ise bir dildeki kelimelerin morfemlerine (kök ve eklerine) ayırarak bunların analizini ve görevlerinin belirlenmesini sağlar. Morfolojik analiz, bir doğal dilin bilgisayar ortamında işlenmesi açısından önemli bir yere sahiptir. Bu yüzden doğru analizinin yapılabilmesi için kapsamlı kelime ve ekler sözlüklerine (lexicon), dilin yazım (ortografik) kurallarına ve bir morfolojik çözümleyiciye ihtiyaç vardır. Biçimbilimsel çözümleme sonucunda kelimelerin sadece kök ve eklerine ayrıştırılması yanı sıra bu kelimelerin türü de (isim, sıfat, fiil, zarf vs.) belirlenmektedir. Kazakça dili de Türkçe gibi

sondan eklemeli bir dildir ve ek bakımından oldukça geniş alana sahiptir. Bu da dilin biçimbilimsel analizini daha karmaşık hale getirmektedir. Dolayısıyla metin içerisinde geçen her kelimenin birden fazla çözümlemesi olabilir ve bu çözümlemelerden hangisinin seçilmesi aşamasına (morfolojik belirsizlik giderme) geçmektedir. Daha sonra ise elde edilen morfolojik çözümleme çıktıları, dilin sentaks (sözdizimsel) analizi için kullanılmaktadır. Aşağıda Şekil 2.1’de, bir dilde verilen bir kelimenin ya da cümlelerin morfolojik ve sözdizimsel analiz aşamaları verilmiştir.



Şekil 2.1: Morfolojik ve sözdizimsel analiz akış diyagramı.

2.3. KAZAKÇA EK ÇEŞİTLERİ

Bu bölümde Kazakça dilinde bulunan ekler (ek grupları) araştırılmıştır. Kazakçada, yapım (derivational) veya çekim (inflectional) yoluyla kelime oluşturma için kullanılan yaygın ekler (Biray ve diğ., 2015) detaylı olarak verilmiştir. Uygulama sırasında eklerin daha iyi kullanılabilmesi için ekler için bir kodlama şeması oluşturulmuştur.

Kodlama Şeması

Biçimbilimsel gösterimde, büyük harfle gösterilenler gramatik kategorisini (türünü) gösterir. Parantezler ise öncesine bağlı olarak ünlü/ünsüz düşmesini/eklenmesini gösterir. Alt çizgi ‘_’ ise kendisinden sonra gelen ekin, eklendiği gövdeden ayrı yazıldığını gösterir.

$P_0P_1P_2_P_3\dots$

Burada;

P_0 : Final (son) gramatik tür (kategori)

P_1 : Kaynak gramatik tür

P_2 : Çekim türü. I çekim (inflection) ve D ise yapım (derivation)

Daha sonra ise ekin adı gelir.

Gramatik türler (kategoriler):

N	İsim
V	Fiil
A	Zarf
J	Sıfat

Bu kodlama şemasına örnek verilecek olursa; VND_abc eki, abc eki eklendiği isim türündeki gövdeyi fiil yapan yapım ekidir.

Nüve morfolojik çözümleyicide dilin çözümlemesinin yapılması için ekler dosyasına ihtiyaç vardır. Çekim eklerinde yapılan kodlama, ekin anlaşılabilmesi için daha açık bir şekilde yazılırken, yapım ekleri kullanılırken yapılan kodlama ise basitçe şu şekildedir:

$P_iT_P_s_suffix$

P_i : İlk gramatik tür

T : Ekin türü. Çekim eki için I, yapım eki için D

P_s : Ek aldıktan sonra dönüştüğü son gramatik tür

suffix: ekin anımsatıcı adı

Ek çeşitlerinin verildiği tablolarda sütunlar, sırasıyla ilgili ekin *Biçimbilimsel Gösterim* i, ekin (Ofлаzer ve diğ., 1994) kodu, Nüve morfolojik çözümleyicisindeki karşılığı, ekin açıklaması ve sırasıyla da ekin *Latin* ve *Kiril* alfabesiyle örnekleri verilmiştir.

2.3.1. İsim Çekim Ekleri (NNI_XXX)

Çekim ekleri sırasıyla aşağıda verilmiştir (Tablo 2.3):

1. İsim gövdesi (N)
2. Çoğul (NNI_PLU)
3. İyelik (NNI_POSSxx)
4. Hal (NNI_XXX)
5. İlgil (NNI_REL)
6. Soru (NNI_QUE)
7. Eşitlik (NNI_EQUAL)

Tablo 2.3: İsim çekim ekleri.

Bişimbilimsel Gösterim	Oflazer ve diğ. (1994) Kodu	NÜVE Kodu	Açıklama	Örnek (Latin)	Örnek (Kiril)
-LAp	NNI_PLU	NI_PLURAL_LAr	Çoğul Eki	balalar,üyeler	балалар, үйлер
-(I)m	NNI_POSS1s	NI_POSSESSIVE_1sg_(I)m	1.Tekil Şahıs İyelik Eki	balam,üyim	балам, үйім
-(I)mIz	NNI_POSS1p	NI_POSSESSIVE_1pl_(I)mIz	1.Çoğul Şahıs İyelik Eki	баламыз, үйіміз	баламыз, үйіміз
-(I)ñ	NNI_POSS2s	NI_POSSESSIVE_2sg_(I)ñ	2.Tekil Şahıs İyelik Eki	balañ, üyiñ	балаң, үйің
-(I)ñIz	NNI_POSS2p	NI_POSSESSIVE_2pl_(I)ñIz	2.Çoğul Şahıs İyelik Eki	balañız, üyiñiz	балаңыз, үйіңіз
-LArIñ	NNI_POSS2p	NI_POSSESSIVE_2pl_IArIñ	2.Tekil Şahıs İyelik Eki (Nezaket)	balalarıñ, üyleriñ	балаларың, үйлерің
-LArIñIz	NNI_POSS2p	NI_POSSESSIVE_2pl_IArIñIz	2.Çoğul Şahıs İyelik Eki	balalarıñız, üyleriñiz	балаларыңыз, үйлеріңіз
-(c)I	NNI_POSS3s	NI_POSSESSIVE_3sg_(s)I	3.Tekil Şahıs İyelik Eki	balası, üyi	баласы, үйі
-LArI	NNI_POSS3p	NI_POSSESSIVE_3pl_LArI	3.Çoğul Şahıs İyelik Eki	balaları, üyleri	балалары, үйлері
-NI	NNI_OBJ	NI_CASE_ACC_NI	Belirtme Hali	balanı, üydi	баланы, үйді
-н	NNI_OBJ3	NI_CASE_ACC_N(3)	Belirtme Hali (3. Tek. Ş. İye. E. sonra)	balasın, üyin	баласын, үйін
0	NNI_NOM		Yalın Hali	bala,üy	бала, үй
-NIñ	NNI_GEN	NI_CASE_GEN_NIn	İlgi Hali	balanın, üydiñ	баланың, үйдің
-GA	NNI_DAT	NI_CASE_DAT_GA	Yönelme Hali	balağa, üyge	балаға, үйге
-DA	NNI_LOC	NI_CASE_LOC_DA	Bulunma Hali	balada, üyde	балада, үйде
-DAн	NNI_ABL	NI_CASE_ABL_DAn	Ayrılma Hali	baladan, üyden	баладан, үйден
-men	NNI_INCM	NI_CASE_INST_men	Vasıta Hali	balamen	баламен
-menen	NNI_INCMN	NI_CASE_INST_menen	Vasıta Hali	közbenen	көзбенен
-GI	NNI_REL	NI_REL_GI	Aitlik Eki	meydandağı, bugingi	мейдандағы, бүгінгі
-_MA	NNI_QUES	QUES_MA	Soru	jaz ba, üy me	жаз ба, үй ме
-DAй	NNI_EQUAL	NI_CASE_EQU_DAY	Eşitlik Hali	baladay, üydey	баладай, үйдей

2.3.2. İsim Yapım Ekleri (NxD_xxx)

Bu bölümde Kazakçada eklendiği kökü isim/sıfat türüne dönüştüren yapım ekleri verilmiştir. Bu ekler; isimden isim yapan (NND_xxxxx) yapım ekleri, sıfattan isim yapan (NJD_xxx) ya da sıfattan sıfat yapan (JJD_xxx) yapım ekleri, fiilden isim yapan (NVD_xxxx) yapım ekleri ve fiilden sıfat yapan (JVD_xxx) yapım ekleridir.

2.3.2.1. İsimden İsim Yapan Yapım Ekleri (NND_xxx)

Eklendiği isim gövdelerinden isim yapan yapım ekleri aşağıdaki tabloda örneklerle birlikte verilmiştir.

Tablo 2.4: İsimden isim yapan yapım ekleri.

Bişimbilimsel Gösterim	Oflazer ve diğ. (1994) Kodu	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-Ab ¹	NND_AV	ND_NOUN_Av	üşev, altav	үшев, алтав
-DAc	NND_DAS	ND_NOUN_DAs	joldas, attas	жолдас, аттас
-LIK	NND_LIK	ND_NOUN_LIK	balalıq, joldastıq	балалық, жолдастық
-LI	NND_LI	ND_NOUN_LI	paydalı, güldi	пайдалы, гүлді
-pAK	NND_RAK	ND_NOUN_rAK	küştirek, jakşıraq	күштірек, жакшырақ
-clz	NND_SIZ	ND_NOUN_slz	balasız, üysiz	баласыз, үйсіз
-шIK	NJD_ŞIK	ND_NOUN_şIK	üyşik, qalaşıq	үйшік, қалашық
-шіл	NND_ŞIL	ND_NOUN_şIl	oyşıl, Türikşil	ойшыл, Түрікшіл
-шAK	NND_ŞAK	ND_NOUN_şAK	boyunşaq, qulınsaq	бойұншақ, құлынсақ
-шAң	NND_ŞAN	ND_NOUN_şAn	kirşeñ, boyşañ	кіршең, бойшаñ
-шA	NND_ŞA	ND_NOUN_şA	kitapşa, bölimşe	кітапша, бөлімше
-ші	NND_ŞI	ND_NOUN_şI	tilşi, okuşı	тілші, оқушы
-тай	NND_TAY	ND_NOUN_DIMINUTIVE_tay	ağatay, erketay	ағатай, еркетай

2.3.2.2. Fiilden İsim Yapan Yapım Ekleri (NVD_xxx)

Eklendiği fiil gövdelerinden isim yapan yapım ekleri örneklerle birlikte aşağıdaki tabloda verilmiştir.

¹ Bu ek birden yediye kadar olan sayılardan sonra gelerek beraberlik anlamı katmaktadır.

Tablo 2.5: Fiilden isim yapan yapım ekleri.

Bişimbilimsel Gösterim	Offlazer ve diğ. (1994) Kodu	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-MA	NVD_MA	VD_NOUN_MA	ilme, bölme	ілме, бөлме
-Im	NVD_IM	VD_NOUN_IM	bölim, kiyim	бөлім, кийім
-Ic	NVD_IS	VD_NOUN_Is	urıs, jüris	ұрыс, жүріс
-GI	NVD_GI	VD_NOUN_GI	sızgı, uyqı	сызгы, ұйқы
-IK	NVD_IK	VD_NOUN_IK	taraq, qazıq	тарақ, қазық
-AK	NVD_AK	VD_NOUN_AK	turaq, uşaq	тұрақ, ұшақ
-IndI	NVD_INDI	VD_NOUN_IndI	asırandı, jasandı	асыранды, жасанды
-GIş	NVD_GIŞ	VD_NOUN_GIş	tutqış, sezgiş	тұтқыш, сезгіш
-Iş	NVD_IŞ	VD_NOUN_Iş	quvanış, süyiniş	құваныш, сүйініш
-In	NVD_IN	VD_NOUN_In	eğın, kelin	еғін, келін
-MAK	NVD_MAK	VD_NOUN_MAK	ilmek, oymaq	ілмек, оймақ
-GIn	NVD_GIN	VD_NOUN_GIn	bosqın, köşkin	босқын, көшкін
-mAl	NVD_MAL	VD_NOUN_mAl	süyemel, tasımal	сүйемел, тасымал
-şAK	NVD_ŞAK	VD_NOUN_şAK	bölşek, jutqınsaq	бөлшек, жұтқынсақ

2.3.2.3. Sıfattan İsim Yapan (NJD_xxx) ya da Sıfattan Sıfat Yapan Yapım Ekleri (JJD_xxx)

Sıfat kök veya gövdelerine gelerek bunları sıfat ya da isim yapan yapım ekleri aşağıdaki tabloda örneklerle birlikte verilmiştir.

Tablo 2.6: Sıfattan isim yapan ya da sıfattan sıfat yapan yapım ekleri.

Bişimbilimsel Gösterim	Offlazer ve diğ. (1994) Kodu	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-LI	NJD_LI	ND_ADJ_LI	altındı, arbalı	алтынды, арбалы
-ҒЫЛТ	NJD_GILT	ND_ADJ_qılt	bozğılt, sarğılt	бозғылт, сарғылт
-DAGAn	NJD_DAGAN	NUM_ADV_DAGAn	ondağan, жүzdegen	ондаған, жүздеген
-нші	JJD_NŞI	NUM_ORD_nŞI	ekinşi, onınşı	екінші, оныншы
-NAn	JJD_NAN	NUM_DISTB_NAn	ekiden, altıdan	екіден, алтыдан

2.3.2.4. Fiilden Sıfat Yapan Yapım Eki (JVD_xxx)

Eklendiği fiil gövdelerinden sıfat yapan yapım ekleri örneklerle birlikte aşağıdaki tabloda verilmiştir.

Tablo 2.7: Fiilden sıfat yapan yapım eki.

Biçimbilimsel Gösterim	Oflazer ve diğ. (1994) Kodu	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-GAH	JVD_GAN	VERBAL_ADJECTIVE_GAn	körgen, jazılğan	көрген, жазылған
-Ap	JVD_AR	VERBAL_ADJECTIVE_Ar	aytar, oqır	айтар, оқыр
-MAc	JVD_MAS	VERBAL_ADJECTIVE_MAs	körmes, barmas	көрмес, бармас
-AtIn	JVD_ATIN	VERBAL_ADJECTIVE_AtIn	esetin, qalatın	есетін, қалатын

2.3.3. Fiil Çekim Ekleri (VVI_xxx)

Türkçede olduğu gibi Kazakçada da fiil çekiminin tam olması için zaman ve şahıs eklerinin olması gerekir. Bu bölümde fiil kök veya gövdesine eklenen çekim ekleri Tablo 2.8’de verilmiştir.

1. Fiil Gövdesi (V)
2. Dönüşlülük (VVI_REFX)
3. İşteşlik (VVI_RECP)
4. Ettirgen (VVI_CAUSx)
5. Edilgen (VVI_PASSx)
6. Negatif (VVI_NEG)
7. Zaman (VVI_Txxxxx)
8. Şahıs (VVI_PERSxxxxx)

Tablo 2.8: Fiil çekim ekleri.

Biçimbilimsel Gösterim	Oflazer ve diğ. (1994) Kodu	NÜVE Kodu	Açıklama	Örnek (Latin)	Örnek (Kiril)
-(I)H	VVI_REFX	VD_REFLEX_In	Dönüşlü	körin, oylan	көрін, ойлан
-(I)c	VVI_RECP	VD_RECIP_Is	İşteş/ Birliktelik	aytış, körıs	айтыс, көріс
-Dİp	VVI_CAUSD	VD_CAUS_Dİr	Ettirgen	keltir, bildir	келтір, білдір
-Gİz	VVI_CAUSG	VD_CAUS_Gİz	Ettirgen	ekkiz, kirgiz	еккіз, кіргіз
-(I)t	VVI_CAUST	VD_CAUS_It	Ettirgen	oqıt, ürkit	оқыт, үркіт
-(I)p	VVI_CAUSR	VD_CAUS_Ir	Ettirgen	asır, pisir	асыр, пісір
-(I)л	VVI_PASSL	VD_PASS_II	Pasif	üzil, juvıl	үзіл, жұвыл
-mA	VVI_NEG	FD_NEGATIVE_mA	Olumsuzluk	urma, körme	ұрма, көрме
- емес	VVI_AUXNEGE	VD_AUXILIARY_NEG_emes	Olumsuzluk	jazmak emespin	жазмак емеспін
- жоқ	VVI_AUXNEGJ	VD_AUXILIARY_NEG_joq	Olumsuzluk	kelgen joqrız	келген жоқпыз
-Ap	VVI_TAORST	VI_TENSE_AORST_Ar	Geniş Z.	berermin, aytarmız	берермін, айтармыз

Tablo 2.8 (Devam):

Biçimbilimsel Gösterim	Oflazer ve diğ. (1994) Kodu	NÜVE Kodu	Açıklama	Örnek (Latin)	Örnek (Kiril)
-(I)п_отыр	VVI_TPROGIP	VI_TENSE_CONT_Ip_otır	Şimdiki Z.	jazıp otırsıñ	жазып отырсың
-(I)п_жатыр		VI_TENSE_CONT_Ip_jatır	Şimdiki Z.	körip jatırsıñdar	көріп жатырсыңдар
--(I)п_жүр		VI_TENSE_CONT_Ip_jür	Şimdiki Z.	kelip jürmiz	келіп жүрміз
-(I)п_тұр		VI_TENSE_CONT_Ip_tur	Şimdiki Z.	atap turmın	атап тұрмын
-A	VVI_TPROGA	VI_TENSE_CONT_A	Şimdiki Z.	alasin, isteysiñ	аласын, істейсің
-(U)вдA	VVI_TPROGU	VI_TENSE_CONT_UvdA	Şimdiki Z.	jazuvdamız, istevdemiz	жазұвдамыз, істевдеміз
-DI	VVI_TPASTD	VI_TENSE_PAST_DI	Geçmiş Z.	bastadıñ, aştıñ	бастадың, аштың
-GAn	VVI_TPASTG	VI_TENSE_PAST_GAn	Geçmiş Z.	kelgenbiz, alğansıñdar	келгенбіз, алғансыңдар
-(I)п	VVI_TPASTP	VI_TENSE_PAST_Ip	Geçmiş Z.	köripti, jetipti	көріпті, жетіпті
-Atın	VVI_TPASTA	VI_TENSE_PAST_Atın	Geçmiş Z.	oylaytın, baratınbiz	ойлайтын, баратынбыз
-GAlı_отыр	VVI_TFUTRG	VI_TENSE_FUT_GAlı_otır	Gelecek Z.	kelgeli otırmın	келгелі отырмын
-GAlı_жатыр		VI_TENSE_FUT_GAlı_jatır	Gelecek Z.	barğalı jatırsıñ	барғалы жатырсың
-GAlı_жүр		VI_TENSE_FUT_GAlı_jür	Gelecek Z.	oqıp jürmin	оқып жүрмін
-GAlı_тұр		VI_TENSE_FUT_GAlı_tur	Gelecek Z.	jazıp tursıñ	жазып тұрсың
-MAK	VVI_TFUTR	VI_TENSE_FUT_MAK	Gelecek Z.	almaqsıñ	алмақсың
-MAKшI	VVI_TFUTRM	VI_TENSE_FUT_MAKшI	Gelecek Z.	kezbekşimiz	кезбекшіміз
-GI	VVI_OPTA	VI_MOOD_OPT_GI	İstek	istegisi keledi	істерісі келеді
-GAй	VVI_OPTAY	VI_MOOD_OPT_GAy	İstek	işkeymiz	ішкейміз
-(U)в	VVI_NECESS	VI_MOOD_NECESS_Uv	Gereklilik	jazuvımız kerek	жазұвымыз керек
-сA	VVI_COND	VI_MOOD_COND_sA	Şart	bilsek	білсек
-(A)йін	VVI_IMP1SG	VI_MOOD_IMPR_1SG_AyIn	Emir	bastayım	бастайым
-сін	VVI_IMP3SG	VI_MOOD_IMPR_3SG_sIn	Emir	bastasın	бастасын
-(A)йік	VVI_IMP1PL	VI_MOOD_IMPR_1PL_AyIK	Emir	bastayıq	бастайық
-ңіз	VVI_IMP2PL	VI_MOOD_IMPR_2PL_ñIz	Emir	bastañız	бастаңыз
-ңдАр	VVI_IMP2PL	VI_MOOD_IMPR_2PL_ñdAr	Emir	bastañdar	бастаңдар
-ңіздАр	VVI_IMP2PL	VI_MOOD_IMPR_2PL_ñIzdAr	Emir	bastañızdar	бастаңыздар
-_екен	VVI_PASTP	SECONDTENSE_PASTPRF_eken	Rivayet	kelgen ekenmin	келген екенмін
-_еді	VVI_NARR	SECONDTENSE_NARR_edi	Hikaye	körip edik	көріп едік
-_болса	VII_SCOND	SECONDTENSE_COND_bolsa	Şart	kelgen bolsañ	келген болсаң

Tablo 2.8 (Devam):

Biçimbilimsel Gösterim	Oflazer ve diğ. (1994) Kodu	NÜVE Kodu	Açıklama	Örnek (Latin)	Örnek (Kiril)
- келеди	VVI_AUXOPT	VD_AUXILIARY_OPT_keledi	Yrd. Fiil ²	istegisi keledi	істерісі келеді
- келмейди	VVI_AUXOPT	VD_AUXILIARY_OPT_kelmeydi	Yrd. Fiil	istegim kelmeydi	істегім келмейді
- келет	VVI_AUXOPT	VD_AUXILIARY_kelet	Yrd. Fiil	barğıñ keletin edi	барғың келетін еді
- керек	VVI_AUXNCS	VD_AUXILIARY_NECESS_kerek	Yrd. Fiil	oqivimiz kerek	оқывымыз керек
-Мін	VVI_PERS1sg1	PERS_PNOUN_1sg_mln	1.Tekil Ş. Eki Tip1	kelgenmin	келгенмін
-сің	VVI_PERS2sg1	PERS_PNOUN_2sg_slñ	2.Tekil Ş. Eki Tip1	kelgensiñ	келгенсің
-сіңдАр	VVI_PERS2sg1	PERS_PNOUN_2pl_slñdAr	2.Tekil Ş. Eki Tip1 (Nez.)	kelgensiñder	келгенсіңдер
-Ді	VVI_PERS3s1	PERS_PNOUN_3sg_DI	3.Tekil Şahıs Eki Tip1	kelgendi	келгенді
-Біз	VVI_PERS1pl1	PERS_PNOUN_1pl_Blz	1.Çoğul Ş. Eki Tip1	kelgenbiz	келгенбіз
-сіз	VVI_PERS2pl1	PERS_PNOUN_2pl_slz	2.Çoğul Ş. Eki Tip1	kelgensiz	келгенсіз
-Ді	VVI_PERS3pl1	PERS_PNOUN_3pl_DI	3.Çoğul Ş. Eki Tip1	kelgendi	келгенді
-м	VVI_PERS1sg2	PERS_POSS_1sg_m	1.Tekil Ş. Eki Tip2	kettim	кеттім
-ң	VVI_PERS2sg2	PERS_POSS_2sg_ñ	2.Tekil Ş. Eki Tip2	kettiñ	кеттің
-ңдАр	VVI_PERS2sg2	PERS_POSS_2pl_ñdAr	2.Tekil Ş. Eki Tip2 (Nez.)	kettiñder	кеттіңдер
-к	VVI_PERS1pl2	PERS_POSS_1pl_k	1.Çoğul Ş. Eki Tip2	kettik	кеттік
-ңіз	VVI_PERS2pl2	PERS_POSS_2pl_ñlz	2.Çoğul Ş. Eki Tip2	kettiñiz	кеттіңіз
-ңіздАр	VVI_PERS2pl2	PERS_POSS_2pl_ñlzdAr	2.Çoğul Ş. Eki Tip2 (Nez.)	kettiñizder	кеттіңіздер

² Bu yardımcı fiiller (Yrd. Fiil) istek kipinin veya gereklilik kipinin çekiminde kullanılmaktadır.

2.3.4. Fiil Yapım Ekleri (VxD_xxx)

Bu bölümde eklendiği gövdeyi fiil türüne dönüştüren yapım ekleri verilmiştir: İsimden fiil yapan (VND_xxxxx) veya sıfattan fiil yapan (VJD_xxxxx) yapım ekleri.

2.3.4.1. İsimden Fiil Yapan Yapım Eki (VND_xxx)

Eklendiği isim gövdelerinden fiil yapan yapım ekleri aşağıdaki tabloda bu ekler verilmiştir.

Tablo 2.9: İsimden fiil yapan yapım ekleri.

Bişimbilimsel Gösterim	Oflazer ve diğ. (1994)	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-cH	VND_SIN	ND_VERB_sl	jatsın, azsın	жатсын, азсын
-cl	VND_SI	ND_VERB_sl	bilgensi, batırsı	білгенсі, батырсы
-pA	VND_PA	ND_VERB_rA	barqıra, dürkire	барқыра, дүркіре
-GAp	VND_GAR	ND_VERB_Gar	sugar, basqar	сұғар, басқар
-LAт	VND_LAT	ND_VERB_LAT	tezdet, borandat	тездет, борандат
-LAc	VND_LAS	ND_VERB_LAs	eldes, sırlas	елдес, сырлас
-LAH	VND_LAN	ND_VERB_Lan	aşulan, maldan	ашұлан, малдан
-LA	VND_LA	ND_VERB_LA	közde, ilgerile	көзде, ілгеріле
-IK	VND_IK	ND_VERB_IK	tarıq, keşik	тарық, кешік
-I	VND_I	ND_VERB_I	juvası, kemi	жұвасы, кемі
-DA	VND_DA	ND_VERB_DA	şırılда, bajılда	шырылда, бажылда
-Aй	VND_AY	ND_VERB_Ay	qaray, küşey	қарай, күшей
-(A)p	VND_AR	ND_VERB_Ar	qısqar, esker	қысқар, ескір
-(A)л	VND_AL	ND_VERB_AI	tiril, suyıl	тіріл, сұйыл
-A	VND_A	ND_VERB_A	tune, sina	түне, сына

2.3.4.2. Fiilden Fiil Yapan Yapım Eki (VVD_xxx)

Eklendiği fiil gövdelerinden fiil yapan yapım ekleridir. Aşağıdaki tabloda bu ekler verilmiştir.

Tablo 2.10: Fiilden fiil yapan yapım eki.

Bişimbilimsel Gösterim	Oflazer ve diğ. (1994)	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-Dİp	VVD_DIR	VD_CAUS_Dir	aldır, ayttır	алдыр, айттыр
-Gİz	VVD_GIZ	VD_CAUS_GIz	aytqız, ötkiz	айтқыз, өткіз
-(I)p	VVD_IP	VD_CAUS_Ir	qaytar, köşir-	қайтар, көшір-
-(I)т	VVD_IT	VD_CAUS_It	küret, azayt	күрет, азайт
-(I)л	VVD_IL	VD_PASS_Il	jıynal, şeşil	жыйнал, шешіл
-GİlA	VVD_GILA	VD_CONT_GIIA	şapqıla, tepkile	шапқыла, тепкіле
-(I)н	VVD_IN	VD_REFLEX_In	juvin, körin	жұвын, көрін
-(I)c	VVD_IS	VD_RECIP_Is	karas, kelis	карас, келіс

2.3.4.3. Sıfattan Zarf Yapan Yapım Eki (AJD_xxx)

İsim ya da sıfat gövdelerine eklenerek, bu gövdeleri zarf yapan eklerdir.

Tablo 2.11: Sıfattan zarf yapan yapım eki.

Bıçimbilimsel Gösterim	Oflazer ve diğ. (1994)	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-DAGAn	AJD_DAGAN	NUM_ADV_DAGAn	jüzdegen, qırıқтаған	жүздеген, қырықтаған

2.3.4.4. Fiilden Zarf Yapan Yapım ekleri (AVD_xxx)

Fiilden zarf yapan yapım ekleri aşağıdaki tabloda verilmiştir.

Tablo 2.12: Fiilden zarf yapan yapım ekleri.

Bıçimbilimsel Gösterim	Oflazer ve diğ. (1994)	NÜVE Kodu	Örnek (Latin)	Örnek (Kiril)
-A ³	AVD_A	VERBAL_ADVERB_A	kele kele	келе келе
-МАй	AVD_MAY	VERBAL_ADVERB_MAY	kelmey, toymay	келмей, тоймай
-GAndA	AVD_GANDA	VERBAL_ADVERB_GAndA	kelgende, aytqanda	келгенде, айтқанда
-GAnşA	AVD_GANŞA	VERBAL_ADVERB_GAnşA	aytqanşa, kelgenşe	айтқанша, келгенше
-GAlı	AVD_GALI	VERBAL_ADVERB_GAlı	qaytqalı, körmegeli	қайтқалы, көрмегелі
-MAstAn	AVD_MASTAN	VERBAL_ADVERB_MAstAn	bilmesten, körmesten	білместен, көрместен
-MAyInşA	AVD_MAYINŞA	VERBAL_ADVERB_MAYInşA	aytpayınşa, oylamayınşa	айтпайынша, ойламайынша
-(I)п	AVD_IP	VERBAL_ADVERB_Ip	körüp, kelip	көрүп, келіп

³ Bu ek genellikle birleşik fiil oluşumunda ve tekrar gruplarında kullanılır.

3. MALZEME VE YÖNTEM

3.1. KAZAKÇA BİÇİMBİLİMSEL ANALİZ

3.1.1. Biçimbilimsel Çözümleme

Biçimbilimsel çalışmaların temelindeki yapı taşı sözcüktür. Doğal dil işlemenin de en temel seviyesi biçimbilimsel çözümlemedir. Biçimbilimsel çözümleme, kelime türünün belirlenmesi, anlamsal ve sözdizimsel çözümleme, makine çevirisi sistemleri açısından önemli bir yere sahiptir. Biçimbilimsel çözümleme verilen bir sözcüğü yapısını bilgisayar ortamında otomatik olarak çözümlenmesini sağlamaktadır. Böylece sözcük en küçük anlamlı birim olan eklere (morfemlere) ayrıştırılmaktadır. Örneğin “evler” kelimesinin biçimbilimsel çözümlemesi sonucunda “ev+ler” ya da “ev/isim+çoğul” şeklinde bir çözüm elde edilir. Verilen bir kelime için çözüm sayısı birden fazla olabilir.

Doğal dillerle yapılan birçok çalışmaların temelinde 2 yaklaşım vardır: kural tabanlı yaklaşımlar ve istatistiksel tabanlı yaklaşımlar (Eryiğit, 2012). İki yaklaşım türünde de öncelikle dilin yapısıyla ilgili bilgiler bilgisayar ortamına aktarılır. Daha sonra aktarılan bilgiler, doğal dil verileri üzerinde uygulanarak sonuçlar elde edilir. Doğal dillerin biçimbilimsel yapısı incelendiğinde sistematik bir yapı olduğu ve bu yapının belirli kurallar dâhilinde olduğu gözlemlenmektedir. Bu yüzden birçok biçimbilimsel çözümleme çalışmaları kural tabanlı yaklaşımı benimsemektedir.

Kural tabanlı biçimbilimsel çözümleme yaklaşımları yukarıdan-aşağıya (top-down) veya aşağıdan-yukarıya (bottom-up) çözümleme şeklindedir. Literatürde bazı çalışmalarda bu iki yaklaşımdan faydalandığı görülmektedir. Bu iki yaklaşım dışında FST (Finite State Transducer) tabanlı biçimbilimsel çözümleyiciler de bulunmaktadır.

Yukarıdan-aşağıya çözümlemelerde verilen sözcük için çözüm bulunabilmesi için sözcük üzerindeki ekler tek tek bulunup kurallara uygunluğu kontrol edilir. Aşağıdan-yukarıya çözümlemelerde ise bir gövdeye eklenebilecek olası tüm ekler eklenerek oluşturulan yeni sözcüğün incelenerek istenilen sözcükle eşleşip eşleşmediği kontrol edilir.

Top-down çözümleyicilere Eryiğit ve Adalı (2004) geliştirmiş olduğu çözümleyici, bottom-up çözümleyicilere, Akin ve Akin (2007)'in geliştirmiş olduğu Zemberek, FST tabanlıya, Çöltekin (2010) ve Çöltekin (2013)'in Türkçe için yapmış oldukları çalışmalar verilebilir.

Dilmaç makina çevirisi sistemi (Shylov, 2008) Türki diller için geliştirilen sayısı çok az morfolojik çözümleyicilerden biridir. Nüve (Zafer, 2015), başta Türkçe olmak üzere Türki diller için geliştirilen top-down (yukarıdan-aşağıya) morfolojik çözümleyici ve makina çevirisi kütüphanesidir.

Biçimbilimsel çözümleme konusunda yapılan ilk çalışmalar sonlu durumlu dönüştürücüler (Finite State Transducer, FST) tabanlı olduğu görülmektedir. FST tabanlı yaklaşımlarda dilin yapısını tanımlayan biçimbilimsel kurallar, bağlama duyarlı yazım kurallarıyla belirtilmektedir. Bu kurallar hazırlanarak bir sonlu durum dönüştürücüde tanımlanmaktadır. Daha sonra dönüştürücüye verilen girdi üzerinde bu kurallar sırayla uygulanarak sonuç üretilir. Bu yaklaşımdaki önemli nokta dönüştürücüye verilen girdidir. Bu kurallar bu girdiler üzerinde uygulandığı için, girdiye bağlı bir kısıtlama görülmektedir. Yani bir kural kendinden önceki kuralın çıktısı üzerinde değişiklik yapamamaktadır. Sonlu durum dönüştürücülerin bu sorunu Koskenniemi (1983)'nin İki Düzeyli Biçimbilim (Two Level Morphology) yöntemiyle çözülmüştür. Bir dilin biçimbilimsel çözümlemesinde kullanılan yaygın yöntem olan bu yöntem başta Türkçe olmak üzere Fince, Korece, Türkmençe, Uygurca gibi bitişken dillerin biçimbilimsel çözümü için kullanılmıştır. Bunun yanı sıra İngilizce, Japonca gibi farklı biçimbilimsel özelliklere sahip dillerin çözümlenmesi için de kullanıldığı görülmüştür.

Kazakça için morfolojik analiz yapan çözümleyici sayıları oldukça azdır. Kessikbayeva ve Cicekli (2014) Xerox Finite State Tool kullanarak FST tabanlı iki düzeyli morfolojik analizini yapmıştır. Bu yaklaşımında kurallar hazırlanarak dilin analizi sağlanmıştır. Rastgele haberlerden toplanan 1000 kelimelik (daha önce veri setinde geçmemiş) veri setini kullanarak yaklaşımın başarısı yaklaşık olarak %96 olarak ölçülmüştür. Bekmanova ve diğ. (2017), Kazakça morfolojik özelliklerden (features) yararlanarak uniform symbol system (tek tip sembol sistemi) oluşturulmuştur. Bu sistemle Kazakça morfolojik analiz yapılmıştır.

(Makhambetov ve diğ., 2014, Makhambetov ve diğ., 2015) Kazakça için data-driven yaklaşımını kullanarak morfolojik analizini yapmıştır. Bu çalışmalarda morfolojik analiz yapılırken kuralların yazılmasına ihtiyaç duyulmadan kullanılan veri setinden yola çıkarak

Kazakça hem çekim hem de yapım (inflectional and derivational) morfolojisi ele alınmıştır. Yaklaşımlarda *transition label* kullanılarak ekler etiketlenmiştir. Yaklaşım iki aşamadan oluşmaktadır: verilen kelimeyi morphemlerine (eklerine) ayırma (segmenting) ve daha sonra morphemler arasında en olası olanı seçme. Segmenting (bölümleme) aşamasında rekürsif (recursive) fonksiyon kullanılarak soldan sağa kelime ayrıştırılır. Ayrıştırma işlemi sol tarafta sözlük içerisinde geçen bir kelime veya gövde buluncaya kadar devam eder. Bu yaklaşım kullanılan veri setinde %90 başarı sağlamasına rağmen, elde edilen hataların, yaklaşımın içerikten bağımsız olmasından kaynaklandığını savunmuşlar.

3.1.2. İki Düzeyli Morfoloji

Birçok dilin morfolojisinin tanımlanmasında kullanılan İki Düzeyli Morfoloji yöntemi bir dilin biçimbilimini sonlu durum makinalarıyla ifade edilmesini kapsar. 2 adımdan oluşur:

- Yazım kuralları iki düzeyli kurallar şeklinde tanımlanır.
- Çekim ve yapım eklerinin sıralanması (morfotaktik) Sonlu Durum Makinası (SDM) kullanılarak tanımlanır.

Bu yöntem içerisinde bir sözcüğün iki gösterimi vardır: yapısal (lexical) biçim ve yüzeysel biçim (surface). Yapısal biçim, bir sözcüğün hangi biçimbilimsel yapılardan oluştuğunun yani sözcüğe eklenmiş çekim veya yapım eklerinin özellikleri ve türleri, eklenme sırası gibi özelliklerin belirtildiği biçimdir. Yüzeysel biçim de sözcüğün çekim ve/veya yapım eklerini aldıktan sonra metin içerisinde görünen yazımsal biçimdir.

Bu iki düzeyli biçimbilimsel yaklaşımda hem yazım kuralları iki düzeyli kurallar olarak tanımlanır hem de yapımsal/çekimsel (derivational/inflectional) biçim dizgeleri, gövdelere hangi eklerin ve bu eklerin hangi sırada gelebileceğini gösteren dizgeler, sonlu durum makinaları (SDM) ile tanımlanır.

Yapısal ve yüzeysel biçimleri arasındaki dönüşümler yazım kurallarına (orthographic rules) bağlı olarak gerçekleşir. Örneğin, Kazakçadaki “kitap” sözcüğü ünlüyle başlayan “(I)m” 1.tekil şahıs iyelik ekini aldığı anda “p” harfi yumuşayarak “b” harfine dönüşür. İyelik ekinin ilk harfi olan I meta harfi (birden fazla harfi temsil eden harf kümesi) önceki kelimenin son ünlüsüne bağlı olarak “ı” harfine dönüşür. Ayrıca ekin lexical (yapısal) gösteriminde I meta harfi parantez içerisinde gösterilmesinin sebebi ise önceki kelimenin son harfine bağlı olarak ünlü düşmesine

uğrayacağını göstermektedir. Özetleyecek olursak “kitap+(I)m” yapısal gösterimdeki bu kelime belirli ortografik kurallar çerçevesinde yüzeysel gösterimde “kitabım” haline dönüşür.

Yapısal ve yüzeysel gösterim arasındaki değişimler 4 temel kuralla gösterilebilir. Burada, a ve b bir fonemin sırasıyla yapısal ve yüzeysel biçimleridir:

- $a : b \Rightarrow LC_RC$

Yapısal a nın kendinden önce ve sonra belirtilen bağlamları sağladığında yüzeysel b ye dönüştüğünü gösterir. Bu dönüşüm her zaman gerçekleşmeyebilir.

- $a : b \leq LC_RC$

Yapısal biçimdeki a fonemi, kendinden önce ve sonra belirtilen bağlamlar sağlandığında yüzeysel biçimde b ye mutlaka dönüşür.

- $a : b \Leftrightarrow LC_RC$

Yapısal biçimdeki bir a fonemi, kendinden önce ve sonra belirtilen bağlamlar sağlandığında yüzeysel biçimde b ye mutlaka dönüşür. Sadece bu şartlar sağlandığında bu dönüşüm gerçekleşir.

- $a : b \not\leq LC_RC$

Yapısal biçimdeki bir a fonemi, kendinden önce ve sonra belirtilen bağlamlar sağlandığında yüzeysel biçimde b ye kesinlikle dönüşmez.

3.1.3. Kazakçanın Biçimbilimsel Çözümlemesi

Bir dil morfolojik analizinde en önemli adımlardan biri dildeki ekler, eklerin kök veya gövdelere eklenme sırası, bu eklenme sırasında oluşan ortografik kuralların çıkarımıdır.

3.1.3.1. İki Düzeyli Yazım Kuralları

Kazakçanın biçimbilimsel çözümlemesini yapabilmek için Kazakça gramerine uygun olarak iki düzeyli kurallar yazılmıştır. Bu bölümde bu kuralların iki seviyeli gösterimi ve her bir kuralla ilgili örnekler verilmiştir. Kurallar yazılırken iki ya da daha fazla harften oluşan harf kümelerini

göstermek için büyük harflerle gösterilen meta harfler kullanılmıştır. Kullanılan meta harfler Latin-Kiril karşılıklarıyla birlikte aşağıda verilmiştir.

$$\begin{aligned}
 C &= \{b, \text{ç}, d, f, g, \text{ğ}, h, x, j, k, q, l, m, n, \text{ñ}, p, r, s, \text{ş}, t, v, y, z\} \\
 &= \{\text{б}, \text{ч}, \text{д}, \text{ф}, \text{г}, \text{ф}, h, x, \text{ж}, \text{к}, \text{к}, \text{л}, \text{м}, \text{н}, \text{н}, \text{п}, \text{р}, \text{с}, \text{ш}, \text{т}, \text{в}, \text{й}, \text{з}\} \\
 C_{ts} &= \{f, x, h, k, q, p, s, \text{ş}, t, \text{ç}\} = \{\text{ф}, \text{х}, \text{х}, \text{к}, \text{к}, \text{п}, \text{с}, \text{ш}, \text{т}, \text{ч}\} \\
 C_t &= \{b, d, g, \text{ğ}, j, l, m, n, \text{ñ}, r, v, y, z\} = \{\text{б}, \text{д}, \text{г}, \text{г}, \text{ж}, \text{л}, \text{м}, \text{н}, \text{н}, \text{р}, \text{в}, \text{й}, \text{з}\} \\
 V &= \{a, e, \text{ı}, i, o, \text{ö}, u, \text{ü}, \text{ä}\} = \{a, e, \text{ы}, i, o, e, \text{ү}, \text{ү}, \text{ə}\} \\
 V_b &= \{a, \text{ı}, o, u\} = \{a, \text{ы}, o, \text{ү}\} \\
 V_f &= \{e, i, \text{ö}, \text{ü}\} = \{e, \text{и}, e, \text{ү}\} \\
 I &= \{\text{ı}, i\} = \{\text{ы}, i\} \\
 A &= \{a, e\} = \{a, e\} \\
 U &= \{u, \text{ü}\} = \{\text{ү}, \text{ү}\} \\
 K &= \{q, k\} = \{\text{к}, \text{к}\} \\
 N &= \{n, d, t\} = \{\text{н}, \text{д}, \text{т}\} \\
 M &= \{m, b, p\} = \{\text{м}, \text{б}, \text{п}\} \\
 G &= \{g, k, q, \text{ğ}\} = \{\text{г}, \text{к}, \text{к}, \text{ф}\} \\
 D &= \{d, t\} = \{\text{д}, \text{т}\} \\
 L &= \{l, d, t\} = \{\text{л}, \text{д}, \text{т}\}
 \end{aligned}$$

Verilen örneklerde (soldan sağa) 1.sütunda örneğin Kiril alfabesiyle yazımı, 2. sütunda Latin harfleriyle yazımı verilmiştir. 3. sütunda (en sağdaki) ise verilen örneğin kök ve aldığı eklerin tanımını gösterir. Fiil kökleri için V, isim kökleri için N, Zamir kökleri için P, ekler için ise ekler için verilen tablodaki kodlar kullanılmıştır. İtalik olarak yazılanlar ise o sözcüğün Türkçe anlam karşılığıdır.

İlk satırlarda bir sözcüğün yapısal biçimi, ikinci satırda ara (intermediate) biçimi (burada kurallar uygulanmış olup yüzeyde görünmeyen harfler 0 olarak gösterilmiştir.) ve üçüncü satırda bu sözcüğün bir metin içerisinde görünen yüzeysel biçimi gösterilmiştir.

Kuralların yazılırken kullanılan özel simgeler şunlardır:

- * : 0 ya da daha fazla
- ? : 0 ya da 1 kez
- 0 : yüzeyde görünmeyen ses
- @ : kelime sonu

+ : Ek sınırı

[a|b|c] : a veya b veya c

1. Bu kural Kazakça dilinde ünlü uyumu kurallarından biridir. Bir kelimeye eklenen ekteki ünlü, kelimenin son ünlüsüne göre uyum sağlamaktadır. Türkçede olduğu gibi kalın ünlüleri kalın ünlüler, ince ünlüleri ise ince ünlüler takip eder. Yapısal biçimindeki A yüzeysel formda, önceki ünlü kalın ise a, ince ise e olarak görülür.

A:a => V:V_b C* + C*__
A:e => V:V_f C* + C*__

ата+LAp	ata+LAr	N(ata)+NNI_PLU
ата0лар	ata0lar	atalar
аталар	atalar	
жер+DA	jer+DA	N(jer)+NNI_LOC
жер0де	jer0de	yerde
жерде	jerde	

Bu kural şimdiki zaman eki +A, zarf fiil eki +A, sıfat fiil eki +AtIn (+AtIn) gibi bazı özel eklerde bulunan A meta harfinin yüzeysel gösterimdeki değişimlerini sağlamamaktadır. Bu yüzden bu özel durum için bu kural aşağıdaki gibi çalışmaktadır.

A:a => V:V_b C + C*__
A:e => V:V_f C + C*__
A:й => V + __

тұт+A+MIn	tut+A+MIn	V(tut)+VVI_TPROGA+VVI_PERS1sg1
тұт0a0мын	tut0a0mın	tutacağым
тұтамын	tutamın	
оқы+AtIn	oqı+AtIn	V(oku)+JVD_ATIN
оқы0йтын	oqı0ytın	okunacak (yazı)
оқыйтын	oqıytın	
кел+AtIn	kel+AtIn	V(gel)+JVD_ATIN
кел0етін	kel0etin	gelen
келетін	keletin	

Bu kural lexical biçimdeki I meta harfinin yüzeyde önceki ünlünün kalınlık incelik durumuna göre ы (ı) veya і (i) dönüşümünü göstermektedir.

I:ы => V:V_b C* + C*__
I:і => V:V_f C* + C*__

бала+NIң	bala+NIñ	N (çocuk) +NNI_OBJ
бала0ның	bala0nıñ	çocuğun
баланың	balanıñ	
көз+LIK	köz+LIK	N (göz) +NND_LIK
көз0дік	köz0dik	gözlük
көздік	közdik	

2. Bu kural, U meta harfinin Kazakça ünlü uyumuna göre yüzeyde $\underline{y}$ (u) veya $\underline{y}$ (ü) ye değişimi göstermektedir. Son ünlüsü kalın olan gövdeler U meta harfini içeren bir ek alırsa U meta harfi u ya, son ünlüsü ince ise U meta harfi yüzeyde ü ye dönüşür.

$$U:\underline{y} \Rightarrow V:V_{bc}^* + C^* \underline{\quad}$$

$$U:\underline{y} \Rightarrow V:V_{fc}^* + C^* \underline{\quad}$$

жый+Uв	jıy+Uv	V (topla) +JVD_GAN
жый0үв	jıy0uv	toplama
жыйүв	jıyuv	
жет+Uв	jet+Uv	V (var) +JVD_GAN
жет0үв	jet0üv	varma
жетүв	jetüv	

3. Son sesi ünlü, C_t veya C_{ts} olan bir gövde L meta harfini içeren bir ek alırsa, L meta harfi yüzeyde sırasıyla л (l), д (d), т (t) olarak görünür.

$$L:\text{л} \Leftrightarrow V + \underline{\quad}$$

$$L:\text{д} \Leftrightarrow C_t + \underline{\quad}$$

$$L:\text{т} \Leftrightarrow C_{ts} + \underline{\quad}$$

ақыл+LI	aqıl+LI	N (zeka) +JND_LI
ақыл0ды	aqıl0dı	zekalı (zeki)
ақылды	aqıldı	
пайда+LI	payda+LI	N (yarar) +JND_LI
пайда0лы	payda0lı	yararlı
пайдалы	paydalı	
таш+LIk	taş+LIk	N (taş) +NND_LIK
таш0тык	taş0tık	taşlık
таштык	taştık	

4. Son sesi ünlü, C_t veya C_{ts} olan bir gövde N meta harfini içeren bir ek alırsa, N meta harfi yüzeyde sırasıyla н (n), д (d), т (t) olarak görünür.

$N:Н \Leftrightarrow V + \underline{\quad}$
 $N:д \Leftrightarrow C_t + \underline{\quad}$
 $N:т \Leftrightarrow C_{ts} + \underline{\quad}$

бала+NI	bala+NI	N(kid)+NNI_OBJ
бала0ны	bala0nı	<i>çocuğu</i>
баланы	balanı	

көз+NI	köz+NI	N(göz)+NNI_OBJ
көз0дү	köz0dü	<i>gözü</i>
көздү	közdü	

иш+NIн	iş+Nın	N(iş)+NNI_GEN
иш0тин	iş0tin	<i>işin</i>
иштин	iştin	

5. Bu kural D meta harfinin yüzeyde д (d) veya т (t) ye dönüşümünü göstermektedir. D meta harfi sonu sert ünsüzle biten gövdelerden sonra geldiğinde yüzeyde t olarak, diğer durumlarda ise d olarak görülmektedir.

$D:д \Leftrightarrow [V|C_t] + \underline{\quad}$
 $D:т \Leftrightarrow C_{ts} + \underline{\quad}$

көл+DA	köl+DA	N(göl)+NNI_LOC
көл0де	köl0de	<i>gölde</i>
көлде	kölde	

ағаш+DA	ağaş+DA	N(ağaç)+NNI_LOC
ағаш0та	ağaş0ta	<i>ağaçta</i>
ағашта	ağaşta	

6. Kazakçanın iki düzeyli morfolojisini tanımlarken kullanılan bir diğer meta harf M, yüzeyde м (m), б (b) veya п (p) ye dönüşür. Bu dönüşüm tüm durumlarda gerçekleşmediği de görülebilmektedir. Bu yüzden M meta harfi için 3 farklı durum tanımlanmıştır. Eğer sonu sert ünsüzle (C_{ts}) biten bir gövde M meta harfini içeren bir ek alırsa, M meta harfi yüzeyde p ye; eğer gövdenin son sesi э (z) veya с (s) ise M meta harfi yüzeyde b ye, diğer durumlarda ise M meta harfi yüzeyde m ye dönüşür.

$M:м \Rightarrow [V|C_t] - \{э, с\} + \underline{\quad}$
 $M:б \Rightarrow \{э, с\} + \underline{\quad}$
 $M:п \Rightarrow C_{ts} + \underline{\quad}$

қалам+Мен	qalam+Men	N(kalem)+NNI_INCM
қалам0мен	qalam0men	<i>kalemle</i>
қаламмен	qalammen	

жүз+МАЙ	jüz+MAy	V(yüz)+AVD_MAY
жүз0бей	jüz0bey	yüzmeden
жүзбей	jüzbey	
қайт+МАЙІншА	qayt+MAyInşA	V(dön)+AVD_MAYINSA
қайт0пайынша	qayt0payınşa	dönmeyince
қайтпайынша	qaytpaşınca	

M meta harfi için yukarda tanımlanan iki düzeyli kural Kazakça soru eki +MA için doğru sonuç vermediği görülmüştür. Soru ekinin başında M meta harfi eklendiği gövdenin son sesi sert ünsüz ise p ye; m, n, ñ veya z ise b ye; diğer durumlarda ise m ye dönüşür.

M:м => [V|C_t-{м, н, ң, з}] + ____

M:б => {м, н, ң, з} + ____

M:п => C_{ts} + ____

жел+_МА	jel+_MA	N(yel)+NNI_QUES
жел00ме	jel00me	yel mi
жел00ме	jel me	
жаз+_МА	jaz+_MA	N(yaz)+NNI_QUES
жаз00ба	jaz00ba	yaz mı
жаз ба	jaz ba	
жоқ+_МА	jok+_MA	N(yok)+NNI_QUES
жоқ00па	jok00pa	yok mu
жоқ па	jok pa	

M meta harfiyle ilgili bir diğer durum ise fiil gövdesine gelen isim fiil +MA eki gibi M meta harfini içeren bazı eklerdeki M meta harfi yüzeyde farklı koşullar altında yine m, b veya p ye dönüşür.

M:м => {л, м, р, в, й, н, ң} + ____

M:б => [V|C_t-{л, м, р, в, й, н, ң}] + ____

M:п => C_{ts} + ____

бат+МАК	bat+MAK	V(bat)+NVD_MAK
бат0пақ	bat0paq	batmak
батпақ	batpaq	
жаз+МАстАн	jaz+MAstAn	V(yaz)+AVD_MASTAN
жаз0бастан	jaz0bastan	yazmadan
жазбастан	jazbastan	
қара+МАС	qara+MAS	V(bak)+AVD_MAS
қара0мас	qara0mas	bakmayan
қарамас	qara0mas	

7. Ünlü veya sert sessizle biten gövdelerden sonra gelen B meta harfi, yüzeyde б (b) ye, diğer durumlarda ise п (p) ye dönüşür.

B:б <=> [V|C_t] + ____

B:п <=> C_{ts} + ____

көр+GAn+BIз	kör+GAn+BIз	V(gör)+VVI_TPASTG+VVI_PERS1p11
көр0ген0биз	kör0gen0biz	gördük
көр0генбиз	körgenbiz	

8. G meta harfinin yüzeydeki dönüşümlerini gösteren bu kurala göre, eklendiğin sert ünsüzle biten gövdenin son ünlüsüne göre G meta harfi к (k) veya қ (q) olurken, sert ünsüzle bitmeyen gövdelerden sonra ise yine gövdenin son ünlüsünün kalınlık incelik uyumuna göre G meta harfi г (g) ve ғ (ğ) olarak yüzeyde görülür.

G:ғ <=> [V:V_b|V:V_bC_t*] + ____

G:г <=> [V:V_f|V:V_fC_t*] + ____

G:к <=> V:V_fC_{ts}* + ____

G:қ <=> V:V_bC_{ts}* + ____

жан+GAn	jan+GAn	V(yan)+JVD_GAN
жан0ған	jan0ğan	yanan
жанған	janğan	

көр+GAn	kel+GAn	V(gel)+JVD_GAN
кел0ген	kel0gen	gelen
келген	kelgen	

күт+GAn	küt+GAn	V(gel)+JVD_GAN
күт0кен	küt0qen	bekleyen
күткен	kütqen	

соқ+GAn	soq+GAn	V(es)+JVD_GAN
соқ0қан	soq0qan	esen
соққан	soqqan	

9. Kazakçada görülen ünsüz yumuşamasının bir örneği olan bu kurala göre, sonu k ile biten gövdelere ünlüyle başlayan bir ek geldiğinde, к (k) ünsüzü yüzeyde г (g) ye dönüşür.

к:г => ____ + VC?

тілек+cI	тілек+sI	N(dilek)+NNI_POSS3s
тілег00і	tileg00i	dileği (onun)
тілегі	tilegi	

10. Kelime sonunda п (p) ünsüzü, kendinden sonra gelen ek ünlüyle başlıyorsa yüzeyde б (b) ye dönüşür.

п:б => ___ + VC?

кітап+cI	kitap+sI	N(kitap)+NNI_POSS3s
кітаб00ы	kitab00ı	<i>kitabı (onun)</i>
кітабы	kitabı	

11. Sonu қ (q) ile biten gövdelere ünlüyle başlayan bir ek geldiğinde kelimenin sonunda q, ғ (ğ) ye dönüşür.

қ:ғ => ___ + VC?

құлақ+cI	qulaq+sI	N(kulak)+NNI_POSS3s
құлағ00ы	qulağ00ı	<i>kulağı (onun)</i>
құлағы	qulağı	

12. Kazakçada, Türkçede olduğu gibi her hecede en az bir ünlü bulunmalı ve art arda iki ünlü yan yana gelmez. Buna göre, ünlüyle biten gövdelerden sonra ünlüyle başlayan bir ek geldiğinde, ekin başındaki ünlü düşer, yüzeyde görülmez.

V:0 => V + ___

ата+Im	ata+Im	N(baba)+NNI_POSS1s
ата00м	ata00m	<i>babam</i>
атам	atam	

13. Bu kural ауыз (awız, ağız), мойын (moyın, boyun) gibi Kazakça bazı özel kelimeler ünlüyle başlayan ek alması durumunda, son hecelerinde bulunan ünlünün düşmesi kuralıdır.

V:0 => ___C* + V

мойын+Im	moyın+Im	N(boyun)+NNI_POSS1s
мой0ным	moy0n0ım	<i>boynum</i>
мойным	moynım	

14. Lexical yüzeyde c (s) ile başlayan ek, ünsüzle biten gövdeye eklendiğinde, ekin başındaki c (s) sesi yüzeyde düşer.

c:0 <=> C* + ___I

кітап+cI	kitap+sI	N(kitap)+NNI_POSS3s
кітаб00ы	kitab00ı	<i>kitabı</i>
кітабы	kitabı	

15. K meta harfini içeren bir ek, eklendiği gövdenin son ünlüsüne göre, ince ise к (k), kalın ise қ (q) ye dönüşür.

$K:қ \Rightarrow V:V_bC^* + C^* \underline{\quad}$
 $K:к \Rightarrow V:V_fC^* + C^* \underline{\quad}$

ай+LIK	ay+LIK	N(ay)+NND_LIK
ай0лық	ay0l1q	ayl1k
айлық	ayl1q	

16. Ayrılma hal eki +DAn (+DAn) eki, 1. Tekil, 2. Tekil 3. Tekil ve 3. çoğul şahıs eklerinden sonra geldiğinde, ekin başındaki D meta harfi yüzeyde н (n) olarak görülür.

$D:н \Leftrightarrow Im|I_n|cI|LArI + \underline{\quad}An$

үст+cI+DAn	üst+sI+DAn	N(üst)+NNI_POSS3s+NNI_ABL
үст00i0nen	üst00i0nen	üstünden
үстінен	üstinen	

17. Geniş zaman eki +Ap (+Ar) negatif ekinden sonra gelirse, yüzeyde с (s) olarak görülür.

$Ap:c \Leftrightarrow mA + \underline{\quad}$

жең+mA+Ap	ber+MA+Ar	V(yen)+VVI_NEG+VVI_TAORST
жең0ме00с	jeñ0me00s	yenmez
жеңмес	jeñmes	

18. Kazakçada istek kipi birden fazla şekilde çekimlenebilmektedir. İstek kipinin bu türünde, +GI eki şahıs ekini ve келеді (keledi) kelimesini alır. Bu istek kipi geçmiş zamanın hikâye eki _еді (_edi) aldığı anda, keledi kelimesi yüzeyde келетін (keletin) olarak görülür (Burada kelimenin başında yazılan _ sembolü, o kelimenin yüzeyde ayrı yazıldığını gösterir.).

ал+GI+Im+_келеді+_еді ал+GI+Im+_keledi+_edi V(take)+OPT+1PS+AUX+NARR
 ал0ғы00м00келетін00еді ал0ғ100м00keletin00edi alasım geldi

19. Yönelme hal eki +GA, 3. Tekil ve Çoğul iyelik ekleri haricindeki diğer iyelik eklerinden sonra geldiğinde, ekin başındaki G meta harfi yüzeyde düşer. 3. Tekil ve Çoğul iyelik eklerinden sonra yönelme hal eki +GA, yüzeyde н (n) olarak görülmektedir.

$G:0 \Leftrightarrow Im|I_n|LApI_n|LApI_nI_3 + \underline{\quad}A$
 $G:n \Leftrightarrow cI|LApI + \underline{\quad}A$

өз+Im+GA	öz+Im+GA	N(öz)+NNI_POSS1s+NNI_DAT
өз0ім00е	öz0im00e	özüme (kendime)
өзіме	özime	

жол+LArI+GA	jol+LArI+GA	N(yol)+NNI_POSS3p+NNI_DAT
жолдарына	joldarına	yollarına
жолдарына	joldarına	

1., 2. ve 3. Tekil Şahıs zamirlerinden sonra gelen +GA eki, yüzeyde +GAn olarak görülür.

мен+GA	men+GA	P(ben)+NNI_DAT
ма00ған	ma00ğan	bana
маған	mağan	

20. Kişi zamirleri isimler gibi durum (hal) eklerini alabilirler. Bu kural, 1. ve 2. tekil şahıs zamirlerinin (men, sen), durum ekleri alındığında gerçekleşen dönüşüm için yazılmıştır. 1. ve 2. tekil şahıs zamirleri yönelme hal eki +GA aldığı anda, sırasıyla mağan ve sağan olarak yüzey formları gerçekleşir (+GA eki dönüşümü için ayrı ortografik kuralı yazılmıştır). Vasıta hal ekini +мен (+men) alırlarsa, zamir ve sonek arasında i ünlüsü türer ve menimen, senimen olarak yüzeyde gerçekleşecektir.

сен+GA	sen+GA	P(sen)+NNI_DAT
са00ған	sa00ğan	sana
саған	sağan	

21. 3. Tekil şahıs ол (o) zamirinin son harfi olan л (l) ilgi, yönelme, bulunma, yönelme ve ayrılma hal eklerinden birini aldığı anda düşer. Eğer vasıta hal ekini alırsa l yüzeyde n_l olarak görülür.

ол+NI	ol+NI	P(o)+NNI_ACC
о00ны	o00n _l	onu
оны	on _l	
ол+мен	ol+men	P(o)+NNI_INCM
оны0мен	on _l 0men	onunla
онымен	onimen	

3.1.3.2. Morfotaktik

Mantıksal bir kelimenin elde edilmesi için ekler arasındaki ilişkiyi, eklerin sıralanmasını ve modellenmesini düzenleyen kurallar dizisine morfotaktik denir. Bu kurallar Sonlu Durum Makinaları (SDM) ile gösterilir. SDM, durumlar (states) kümesi ve bu durumlar arasındaki geçişlerden (transitions) oluşan yönlü bir graftır. Bu grafta geçişler, sıralanışı belirli olan çekim veya yapım ekleriyle sağlanır. Başlangıç durumları (state) sözlük içerisinde tanımlı olan sözcük türü isim, fiil, zamir, zarf ya da sıfat vs. olan bir gövdeyi (kökü) temsil eder. Ara durumlar (intermediate states) ise ek alarak oluşan o gövdenin türünün gösterildiği durum olarak tanımlanabilir. Son durum (final state) ise başlangıç durumunda verilen gövdenin ara

durumlardaki geçişlerle eklerin eklenmesiyle çekimlenen gövdenin son halini gösterir. Son durum, iç içe iki çember olarak gösterilir ve kendisinden sonra hiçbir ekin gelmesi mümkün değildir. Durumlar arasında geçişlerinde kullanılan “0” sembolü, o durumlar arasında ek almadan geçişin mümkün olduğunu gösterir.

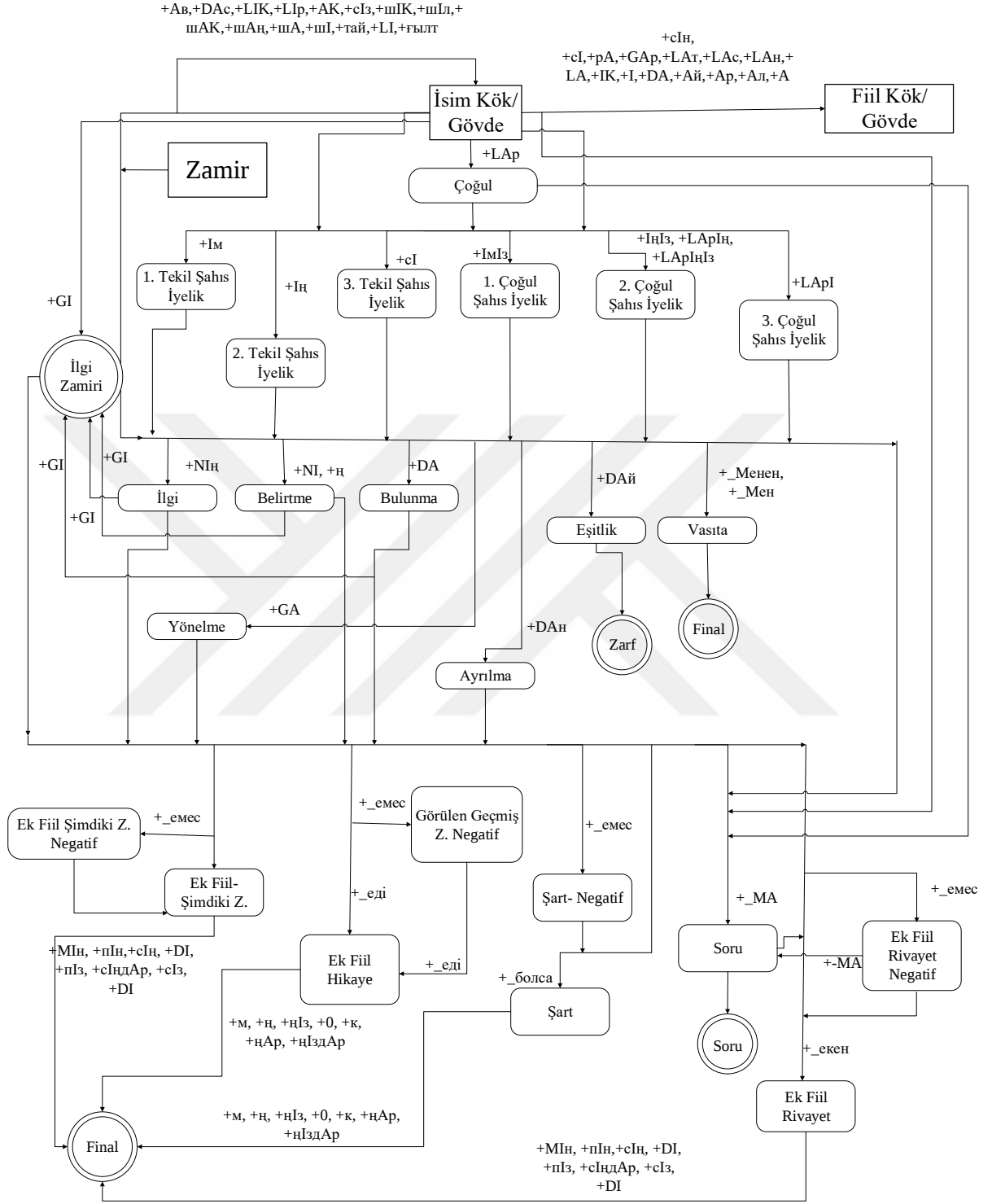
3.1.3.2.1. Nominal Morfotaktik

Kazakça isim soylu gövdeler (isim olan veya yapım eklerini alarak isim türüne dönüşen) ve fiil soylu gövdeler (fiil olan veya yapım eklerini alarak fiil türüne dönüşen) için olmak üzere iki sonlu durum makinası çizilmiştir. Bu sonlu durum makinalarıyla gövdelere gelen çekim eklerinin sırası tanımlanmıştır. İsim ve fiil soylu gövdeler için durum makinaları birbirine bağlıdır. Zamirler ise isimlerde olduğu gibi hal eklerini alarak çekimlenmektedir. Temel nominal model olarak adlandırılan isim çekimleri için basit bir formül aşağıda verilmiştir. Tablo 3.1’de isim kök/gövdelerine gelen çekim ekleri verilmiştir.

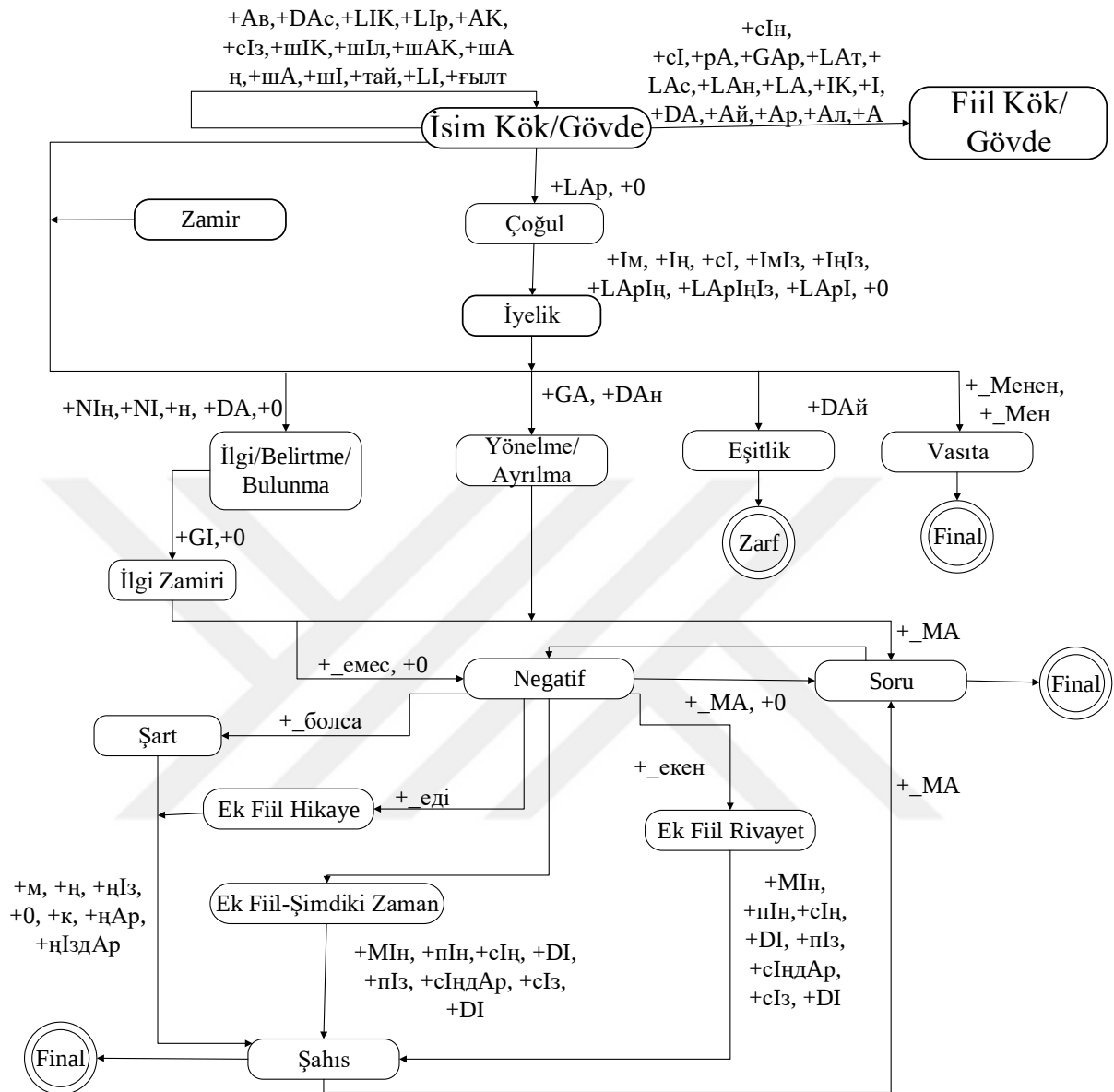
$$\text{İsim} + \text{Çoğul} + \text{İyelik} + \text{Hal} + \text{İlgi}$$

Tablo 3.1: İsim gövdelerine gelen çekim ekleri

Çoğul Eki	+LAp
İyelik Ekleri	+Im, +Iң, +cI, +ImIз, +IңIз, +LApIң, +LApIңIз, +LApI
Hal Ekleri	+NIң, +NI, +H, +GA, +DA, +DAH, +DAй, +MeHeH, +MeH
İlgi Eki	+GI



Şekil 3.1: İsim gövdeleri için SDM modeli (kapsamlı).



Şekil 3.2: İsim gövdeleri için SDM modeli (özet).

İsim kök veya isim soylu gövdeler için çizilmiş SDM modeline bakıldığında (Şekil 3.1’de kapsamlı, Şekil 3.2’de özet olarak verilen), isim gövdesi (Nominal Root) isimden fiil yapan yapım eklerinden (+CИH, +CI, +pA, +GAp, +LAT, +LAC, +LAH, +LA, +IK, +I, +DA, +Aй, +Ap, +Aл, +A) birini alarak fiil gövdeleri için çizilmiş olan SDM’ye bağlanır. Kazakça bir isim kök/gövdesi çoğul ekini (+LAp) alarak Çoğul durumuna (Plural) geçer. Daha sonra aldığı iyelik ekine (+IM, +IH, +CI, +IMIз, +IHИз, +LApIH, +LApIHИз, +LApI) göre ilgili iyelik duruma (Possessive) geçer. Bu durumlar arasında

herhangi bir ek almadan da geçebilmektedir. Bir isim çoğul ekini almadan iyelik ekini alabilir ya da sırasıyla ikisini de alarak bir sonraki duruma geçebilir. İsimler çekimlenirken negatiflik anlamını +_emec (emes, değil) kelimesinden alır. Bu şekilde bir isim çekimlenirken almış olduğu eklere göre durumlar arasında geçişler sağlayarak final durumuna (End) ulaşır.

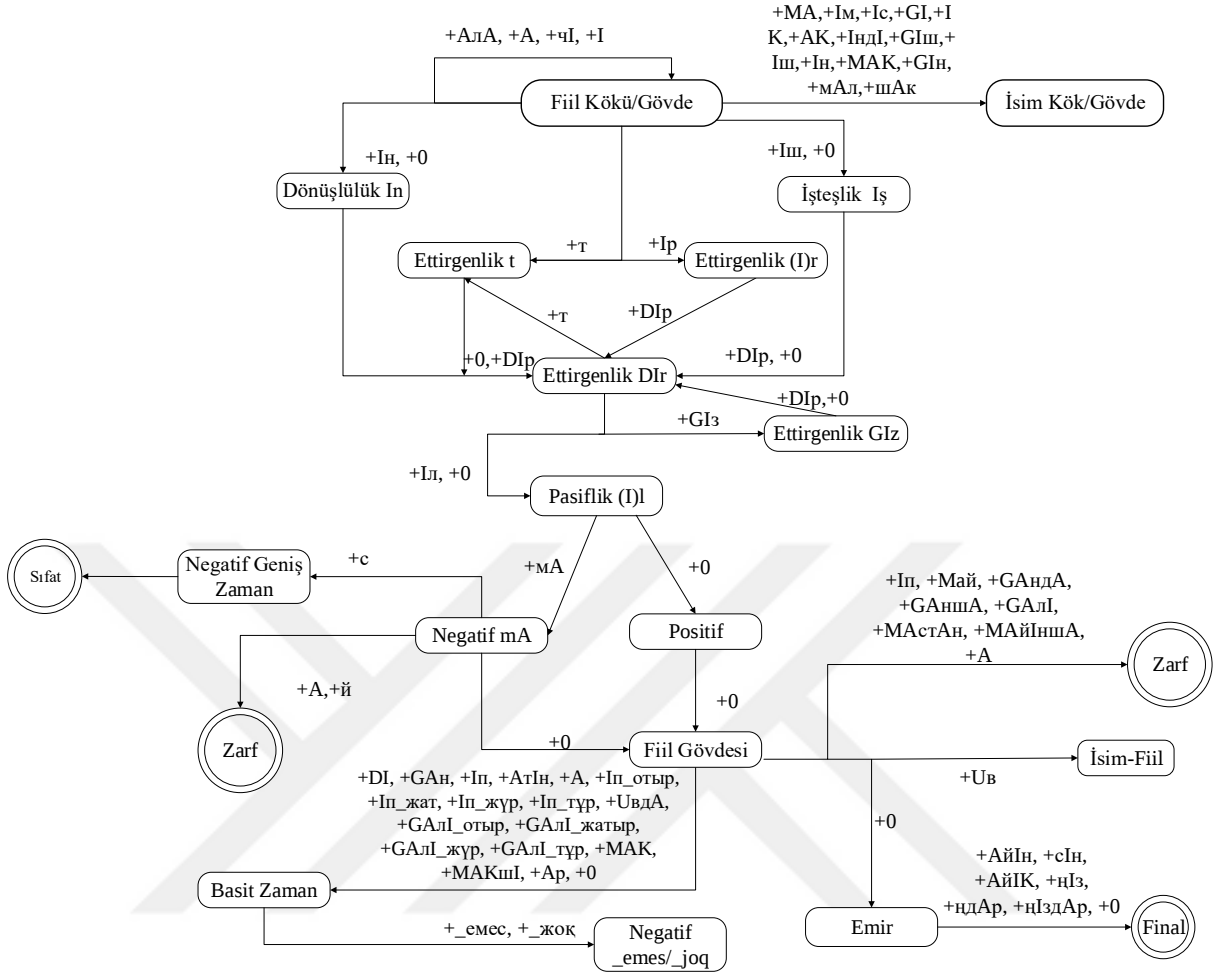
3.1.3.2.2. Verbal Morfotaktik

Kazakçada fiillere gelen eklerin belirli bir sırası vardır. Öncelikle fiil kök veya gövdesine önce fiil çekim ekleri (işteşlik, ettirgenlik, edilgenlik, dönüşlülük vs.), daha sonrasında basit zaman ekleri (geçmiş, gelecek, şimdiki zaman) ya da kip ekleri (istek, gereklilik, şart), sonrasında varsa bileşik zaman ekleri (hikâye, rivayet, şart) gelir. En son olarak şahıs eklerinden biri gelir. Kazakçada iki tip şahıs ekleri bulunmaktadır. Bu şahıs ekleri belirli zamanlara gelir.

Kazakçada fiillere negatiflik anlamı olumsuzluk eki +MA ve yok (yok), emes (değil) kelimeleriyle sağlanır. İki durum arasında fark ise fiil çekimlerinde geliş sırasıdır. +MA olumsuzluk eki zaman ve şahıs eklerinden önce fiil gövdesine eklenirken, yok, emes kelimeleri ise zaman ekleriyle şahıs ekleri arasına gelir. Temel verbal (fiil kök veya gövdesi) model olarak adlandırılan fiil çekimleri için basit bir formül aşağıda verilmiştir.

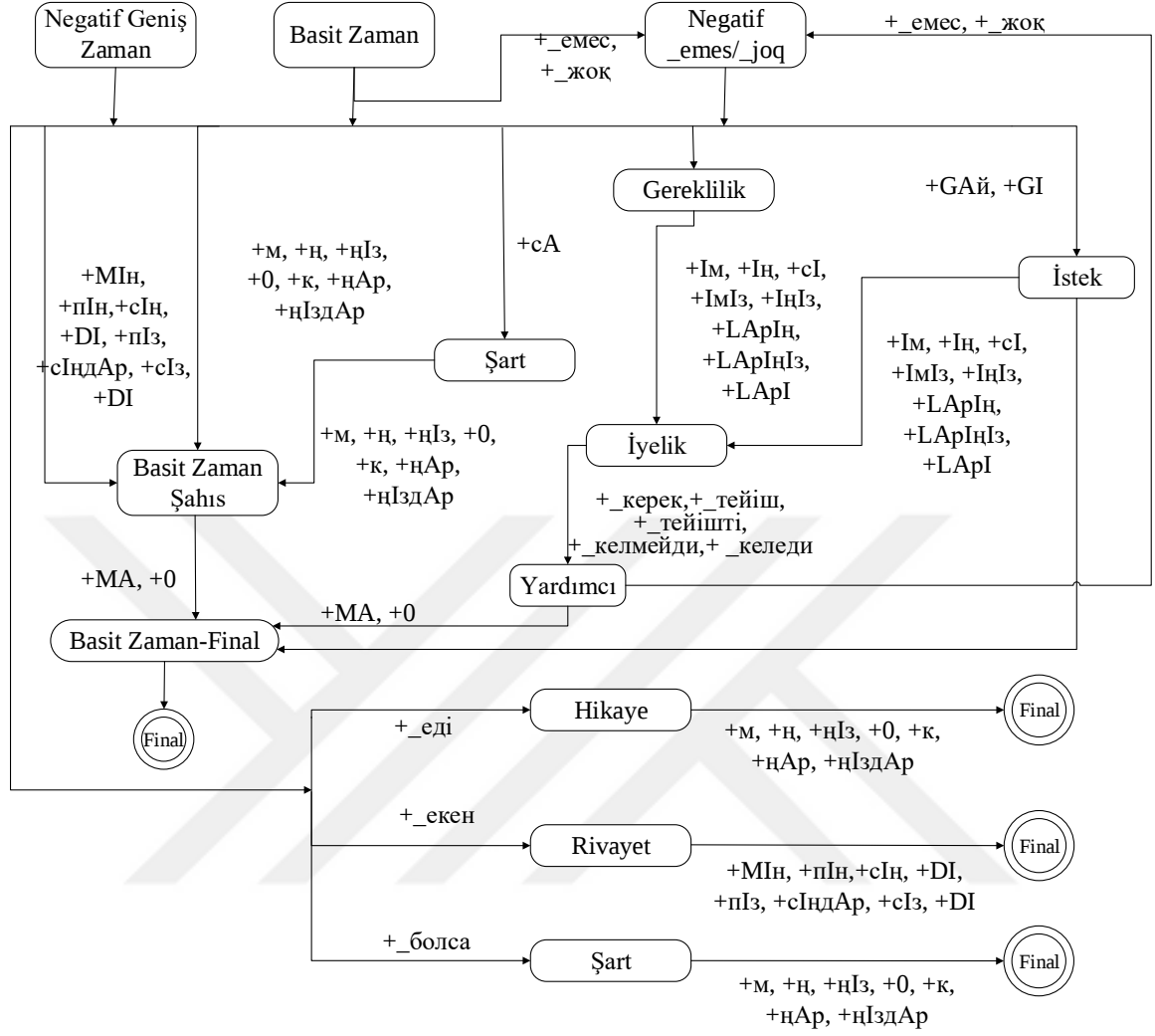
Verb + Negative_MA + Tense + Tense | Mood + Person

Verb + Tense + Negative_jok/emes + Person



Şekil 3.3: Fiil gövdeleri için SDM modeli birinci kısım.

Verbal morfotaktik için çizilen SDM modeli, daha iyi anlaşılması için Şekil 3.3 ve Şekil 3.4’de iki ayrı kısım olarak verilen SDM modelleri birbirine bağlıdır ve birbirinin devamı şeklindedir.



Şekil 3.4: Fiil gövdeleri için SDM modeli ikinci kısım.

3.2. MORFOLOJİK BELİRSİZLİK GİDERME

Bir morfolojik çözümleyicinin verilen bir kelime için birden fazla sonuç bulması, çözümleyici açısından bir avantaj olarak kabul edilmektedir. Fakat bu hangi çözümün seçilmesi gerektiği problemini ortaya koymaktadır. Bazı durumlarda kelime tek başına ele alınarak hangi çözümün daha mantıksal olduğu seçilebilirken, bazı durumlarda ise o kelimenin bulunduğu metin, bir önceki ya da bir sonraki kelime dikkate alınarak seçilmektedir. Örneğin Kazakça бөлме (bölme) kelimesinin Nuve morfolojik çözümleyiciye girdi olarak verildiğinde elde edilen sonuçlar:

1. бөл/VERB + FD_NEGATIVE_mA + VI_MOOD_IMPR_2SG_0
2. бөл/VERB + VD_NOUN_MA

Bu çözümlerden ilkinde бөл (böl) fiili negatiflik ekini almış (ve 2. Tekil şahıs emir kipine göre çekimlenmiştir. Bu ekin yüzeyde gösterimi yoktur.). İkinci çözümde ise бөл (böl) fiili fiilden isim yapan +МА ekini alarak бөлме (bölme, oda) kelimesi türemiştir. Bu бөлме kelimesi tek başına aldığı anda iki çözümde hem anlam hem de yapı bakımından doğrudur. Hangisinin seçileceğine içinde geçmiş olduğu metne bakılarak karar verilebilir. Eğer yargı bildiriyorsa ilk çözüm, cümle içerisinde isim olarak geçiyorsa ikinci çözüm seçilebilir.

Doğal dil işlemede, birden fazla morfolojik çözüm arasında en doğru olanı seçme problemine **morfolojik belirsizlik giderme** (morphological disambiguation) denir. Morfolojik belirsizlik gidermenin girdisi bir morfolojik çözümleyicinin sonuçlarıdır. Morfolojik belirsizlik giderme cümle analizinde, anlam analizinde ve makine çevirisi çalışmalarında önemli bir yere sahiptir. Kazakça, Türkçe gibi ek bakımından zengin olan dillerde morfolojik çözüm sayısı eklere bağlı olarak çok çeşitlidir. Morfolojik belirsizlik giderme yaklaşımları kural tabanlı yaklaşımlar, istatistiksel yaklaşımlar ve bu iki yaklaşımın karışımı olan hibrit yaklaşımlar olarak karşımıza çıkmaktadır.

Kural tabanlı morfolojik belirsizlik giderme yaklaşımlarında elle hazırlanmış kurallar dizisi bulunmaktadır. Bu kurallara bakılarak kelimenin hangi morfolojik çözümünün seçileceğine veya kelime türüne (Part-Of-Speech Tagging, POS) karar verilir. Başta Türkçe olmak üzere sondan eklemeli diller için yapılan çalışmalar Türkçe için (Oflazer ve Tur, 1997, Yuret ve Türe, 2006, Oflazer ve Tur, 1996, Daybelge ve Cicekli, 2007), Uygurca için (Orhun ve diğ., 2010) mevcuttur.

Hakkani-Tür ve diğ. (2002), her bir morfotaktik etiketini (tag) IG (inflectional group) gruplarına ayırarak ele almıştır. Yani her kelime kök ve IG lar dizisinden oluştuğu varsayılmıştır. Daha sonra bu gruplar içerisinde trigram (bir öncesi ve bir sonrası) model uygulanarak istatistiksel hesaplamalar yapmıştır. Bu hesaplamalar sonucunda en yüksek olasılığa sahip tag set seçilmiştir. Türkçede çekim ve yapım bakımından zengin bir morfolojiye sahip olduğundan (büyük etiket seti) dolayı istatistiksel modellerin uygulamasında zorlanmaya sebep olmaktadır.

Morfolojik belirsizlik giderme çalışmaları genellikle bir morfolojik çözümleyicinin çıktısı girdi olarak kabul etmektedir ve morfolojik çözümleme ile belirsizlik giderme ayrı ayrı modellerde çalıştırılmaktadır. Bu çalışmada önerilen Morphnet (Dayanık ve diğ., 2018) hem morfolojik çözümlemenin hem de morfolojik belirsizlik gidermenin tek bir modelde birleşimidir.

Morphnet, seq2seq RNN kullanarak hem çözümlemeyi hem de belirsizlik gidermeyi birleştirilmesi sağlanmıştır. Bu modele taglenmiş veri seti eğitim için verildiğinde, doğru morfolojik analizinin bulunması noktasında iyi başarı sağladığı vurgulanmıştır. Ayrıca bu model, birçok farklı dilde test edilen başarı oranları karşılaştırılmıştır. Ayrıca bu çalışmada, Türkçe için yeni bir veri seti olan TrMor2018 tanıtılmıştır. Bu çalışmadaki model LSTM tabanlıdır. Encoder ve decoder bölümlerinden oluşmaktadır: Word encoder, context encoder, output encoder ve decoder. MorphNet, 2 farklı tek yönlü (unidirectional) LSTM kelime ve output embedding elde etmek kullanırken, bir tane de çift yönlü (bidirectional) kelime için context embedding elde etmek için kullanır. Sonuç olarak harf-harf ayrılmış bir gövde ve akabinde morfolojik özellikler (features) verir. Modelin başlangıç parametreleri başka bir çalışmadan (Glorot ve Bengio, 2010) alınmıştır. Öğrenme katsayısı 1.6 olarak ve epoch sayısı 6 olarak belirlenmiştir. Bu model test edilirken UD (Universal Dependencies) dan alınan veri kümeleri kullanılmıştır. 34673 cümle ve 460663 kelimedenden oluşan bir veri seti oluşturulmuştur (Trmor2018). Bu veri seti yarı-otomatik belirsizliği (semi-automatically disambiguate) giderilmiştir. Veri seti, %80 train veri seti, %10 development veri seti (eğitim aşamasında kullanılmakta, Devset) ve %10 test veri seti olarak bölünmüştür. Türkçe dili için daha önce yapılmış olan etiketleme çalışmalarına göre bu model daha iyi doğruluk oranı vermiştir. %97 den fazla doğruluk (accuracy) sağlamış.

Akyürek ve diğ. (2019) Trmor2018 veri setini ve morfolojik analiz için dilden bağımsız karakter tabanlı encoder-decoder mimarisini (character-based encoder-decoder architecture) önermişler. Bu modeli başta Türkçenin farklı veri setleri olmak üzere birçok dilde test etmişlerdir. Önermiş oldukları derin öğrenme tabanlı Morse modelinde hem kelimeyi eklerine ayırmayı hem de kelimeler için etiketlemeyi tek bir çatı altında sağlamıştır.

Kessikbayeva ve Cicekli (2014), hem Kazakça morfolojik analiz hem de morfolojik belirsizlik giderme üzerine çalışmışlardır. Önce Foma sonlu durum derleyicisi (Hulden, 2009) üzerinde kural tabanlı morfolojik analiz yaptıktan sonra ortaya çıkan morfolojik belirsizlik durumunda da transformation-based morfolojik belirsizlik giderme yapmışlardır. Çalışmanın çıkış noktası Türkçe için yapılan çalışma olup transformation-based (dönüşüm tabanlı) etiketlemeyi kullanmışlar. Bu etiketleme türü Brill tagger ın (Brill, 1995) bir varyasyonudur (burada her kelimeye bir etiket atanması ve önceden tanımlanmış bir dizi kural kullanılarak değiştirilmesinden oluşan dönüşüm tabanlı bir etiketleme süreci vardır.).

Makhambetov ve diğ. (2015) data-driven morfolojik analizini yaptığı çalışmanın ikinci kısmında morfolojik belirsizlik giderme yapmışlardır. Kazakh National Corpus veri setinde (Makhambetov ve diğ., 2013) morfolojik belirsizlik gidermek için HMM (Hidden Markov Model) kullanmışlar.

Morfolojik belirsizlik gidermek için dilden bağımsız derin öğrenme tabanlı yaklaşımı Toleu ve diğ. (2017) Türkçe ve Kazakça dilleri için test etmişler. Bu modelde kelimeler kök, tür ve ekler dizisi şeklinde (root, POS, morpheme chain) üç parçadan oluştuğu varsayılmış. Kelimenin gösterimi (surface) ve morfolojik analizini öğrenen iki bloktan oluşan modelde, gelen metne göre karakter-karakter okuyarak kelimelerin vektörleri oluşturulmaktadır. Daha sonra uygulanan modellerden BiLSTM Kazakça için en iyi sonucu verirken, Türkçede sadece kelime dağarcığının dışında (OOV, out-of-vocabulary) doğruluğunda (accuracy) DNN* (rightmost analysis embeddings kullanan DNN) den daha iyi sonuç vermiştir.

Kazakça belirsizlik gidermek için hibrit bir model öneren Assylbekov ve diğ. (2016), öncelikle kelimeleri ek sınırlarına (Inflectional Group (IG)) ayırır. Morfolojik analiz kısmında kural tabanlı bir yaklaşım kullanılmış olup, analiz sonucunda elde edilen morfolojik çözümler arasında olasılığı en yüksek olan ek dizisine sahip morfolojik çözüm seçilir. Bu yaklaşım, elle hazırlanmış küçük bir veri seti üzerinde test edilmiştir.

3.3. KAZAKÇA SÖZDİZİMSEL ANALİZ

Bir dilin en küçük anlatım birimi olan cümle; bir düşünceyi, bir durumu, bir olayı yargı bildirerek anlatan kelime ya da kelimeler dizisidir. Cümle kurulabilmesi için yargı bildiren bir çekimli isim ya da fiil yeterlidir. Bir cümlede yargı bildiren öge yüklemidir. Cümlelerin ikinci önemli ögesi ise öznedir. Yüklem taşıdığı şahıs ekine bağlı olarak öznenin tekil veya çoğul olması anlaşılabilir. Bir diğer öge olan nesneyse, fiilin geçişli olup olmamasına bağlı olarak cümle içerisinde bulunur. Yer tamlayıcısı ve zarf ise cümlede bulunan yüklem niteliğine göre zorunlu ya da yardımcı öge olarak nitelendirilir. Cümle içerisinde ögelerin dizilişi şöyledir: yüklem genellikle sona bulunur. Belirtisiz nesne ve bazı zarflar, yüklem hemen yanında yer alır. Diğer ögelerin cümle içerisinde yerleri değişkendir. Yüklem yanında bulunan öge, cümle içerisinde vurgulanmak istenen ögedir.

Sözdizimi, kelimelerle veya kelime gruplarıyla oluşturulan cümlelerin nasıl oluşturulacağını tanımlar. Bu kelimeler ve/veya kelime gruplarına **sözdizimsel bölümler (parçalar)** denir.

Sözdizimsel analiz, bir cümleyi sözdizimsel kısımlarına göre analiz etme sürecidir. Bir cümleyi sözdizimsel kısımlarına analiz eden araca **sözdizimsel ayrıştırıcı** denir. Sözdizimi bilgisi, soru yanıtlama, bilgi çıkarma, cümle oluşturma, çeviri vb. gibi birçok gerçek dünya uygulaması için gereklidir. Bu yüzden ayrıştırıcılar önemli yere sahiptir. Örneğin, bir cümleyi kelime kelime kaynak dilden hedef dile çevirmek çoğunlukla yanlış çevirilere neden olabilir, sözcüklerin sırası genellikle hedef dilde farklı olabilmektedir.

Sözdizimsel analiz, cümlelerin yapısal tanımını oluşturabilmek için morfolojik analizin sonuçlarını kullanır. Amacı, arka arkaya gelen kelimeler grubunun ifade ettiği cümlelerin birimlerini tanımlayan bir yapıya dönüştürmektir. Cümle parçalarının doğru tespiti, anlam analizi için yararlı sonuçlar verir.

Tezin bu bölümünde Kazakçanın sözdizimi (sentaksı) ve sözdizimsel analizi üzerine çalışılmıştır.

3.3.1. Context Free Gramer (Bağlamdan Bağımsız Gramer)

Doğal dillerin gramer yapısını tanımlamak için kullanılan en yaygın matematiksel model (Jurafsky & Martin, 2009) olan Context Free Gramer (CFG), Noam Chomsky tarafından doğal dillerde bulunan kelimelerin ve cümle yapılarını tanımlamak için geliştirilmiştir. Literatürde Phase Structure Düzenli ifadelerle (regular expression) ifade edilebilen diller CFG ile de ifade edilebilir.

Bilgisayar biliminde, CFG nin en yaygın notasyonu Backus–Naur Form (BNF) tir. Temel olarak içerikten bağımsız bir gramer G dört özellik içermelidir; S başlangıç sembolü (starting symbol), Σ sonlu karakterler veya değişkenler (terminals), V sonlu olmayan karakterler veya değişkenler (nonterminals), R bağlantılar (relation) dır. Bir gramerin tanımı sırasında kullanılan bu kümeler aşağıdaki şekilde yazılabilir:

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

CFG kurallar dizisinden oluşur. Bu kurallarda sol tarafta nonterminal sembol var iken, sağ tarafta ise terminal veya nonterminal sembol vardır:

$$S \rightarrow aSb$$

$$S \rightarrow ab$$

Burada S nonterminal bir sembol olup, bu gramer örneği sağında ve solunda eşit sayısı **a** ve **b** sembolleri olan karakter katarları (strings) üretir. Örneğin,

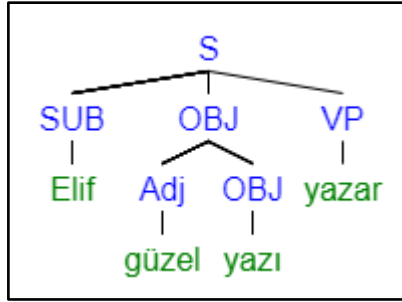
ab, aabb, aaabbb, aaaabbbb, ... şeklinde çoğaltılabilir.

Cümle yapısı ÖNY (ÖZNE-NESNE-YÜKLEM) şeklinde olan Türkçede “Elif”, “güzel”, “yazı”, “yazar” kelimelerinden oluşan temel bir cümle için basit CFG örneği şu şekilde verilebilir:

$S \rightarrow VP$
 $S \rightarrow OBJ VP$
 $S \rightarrow SUB OBJ VP$
 $SUB \rightarrow Elif$
 $Adj \rightarrow güzel$
 $OBJ \rightarrow Adj yazı$
 $OBJ \rightarrow yazı$
 $VP \rightarrow yazar$

Bu gramerle şu cümle üretilir:

“Elif güzel yazı yazar”. Bu cümle için parse (çözüm) ağacı aşağıdaki (Şekil 3.5) gibidir:



Şekil 3.5: "Elif güzel yazı yazar" cümlesi için çözüm (parse) ağacı.

Çözüm ağacını göstermek için kullanılan bir diğer gösterim (parantez) ise;

[S [SUB [Elif]] [OBJ [Adj [güzel]] [OBJ [yazı]]] [VP [yazar]]]

3.3.2. Kazakçanın Gramer Yapısı Bakımından İncelenmesi

- Hem Kazakça hem de Türkçe için cümle türleri yapılarına (basit ve bileşik cümleler), anlamlarına (olumlu, olumsuz, soru), yüklemine göre (kurallı ve devrik), yüklemine türüne (fiil ve isim) göre aynıdır.

- Cümle yapısı iki dilde de özne, nesne, yüklem şeklindedir.
- Kazakçada devrik cümle sıklığı Türkçeye nazaran azdır. Yüklem çoğunlukla hep sonda yer alır.
- +Dİr bildirme ekli Türkçe isim cümleleri Kazakçada yoktur.
- Türkçede *ki* bağlacıyla oluşturulan bileşik cümle türü Kazakçada yoktur.

3.3.3. Kazakça Kelime Grupları ve Sözdizimsel Analizleri

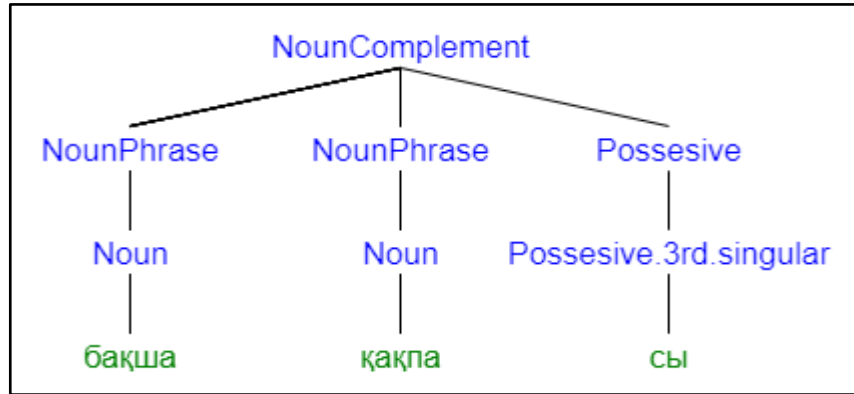
Bu bölümde kelime gruplarının sınıflandırılırken Biray ve diğ. (2015) ve Türkçe için yazılmış olan Karahan (2008) çalışmaları referans alınmıştır. Başlıca Kazakça kelime grupları, tanımlanması için yazılmış olan CFG kuralları, örneği ve çözüm ağaçlarıyla birlikte verilmiştir.

3.3.3.1. İsim Tamlaması (Noun complement)

İsim Tamlaması iki veya daha fazla isim ya da isim soylu parçadan oluşur. İlk parça/element tamlayan, ikincisi ise tamlanandır. Tamlayanın ilgi ekinin aldığı tamlama belirtili isim tamlaması, ek almadığı tamlama ise belirtisiz isim tamlamasıdır. İsim tamlaması örnekleri Şekil 3.6 ve Şekil 3.7’de verilmiştir.

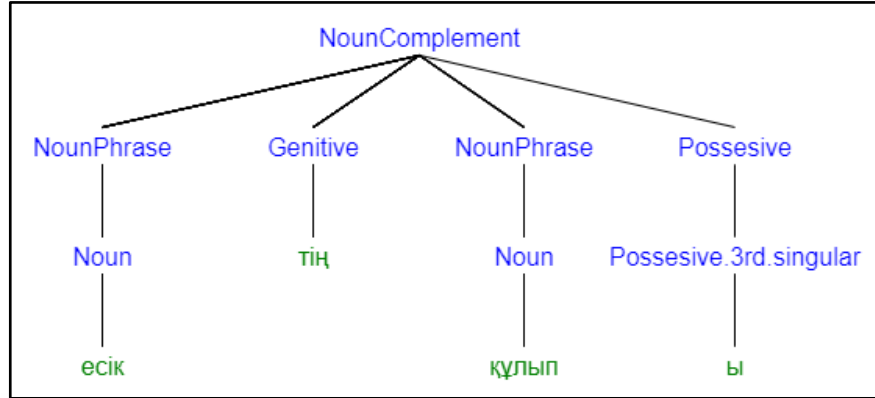
NC → NP NP Poss
 NC → NP gen NP Poss
 Poss → p1s | p2s | p3s | p1p | p2p | p3p

бақша қақпасы
 bahçe kapı+POSS (p3s)
 bahçe kapısı



Şekil 3.6: Belirtisiz isim tamlaması.

есіктің	құлыпы
есік+GEN	құлып+POSS (p3s)
kapının	kilidi



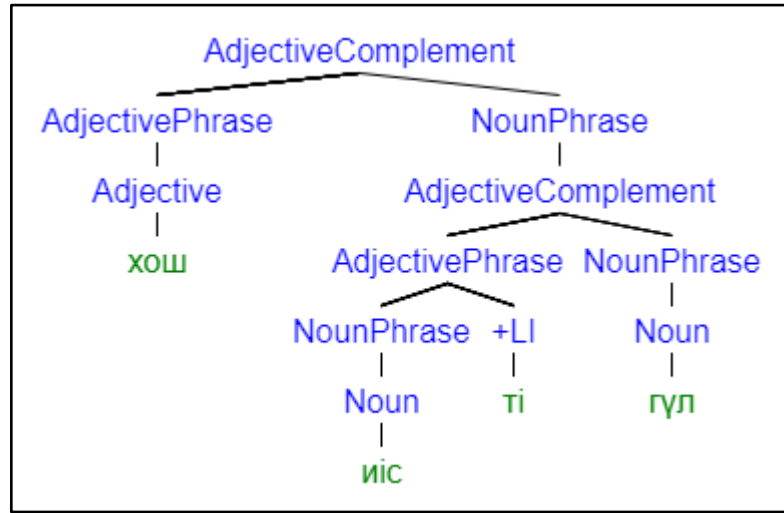
Şekil 3.7: Belirtili isim tamlaması.

3.3.3.2. Sıfat Tamlaması (Adjective Complement)

Sıfat tamlaması, tamlayanın sıfat ya da sıfat grubu olduğu isim tamlamasıdır. Şekil 3.8’de sıfat tamlamasına örnek verilmiştir.

JC	→	JP NP
JC	→	JP cm JC
JC	→	JC cm NP

хош	иісті	гүл
хош	иіс+LI	гүл
hoş	kokulu	çiçek



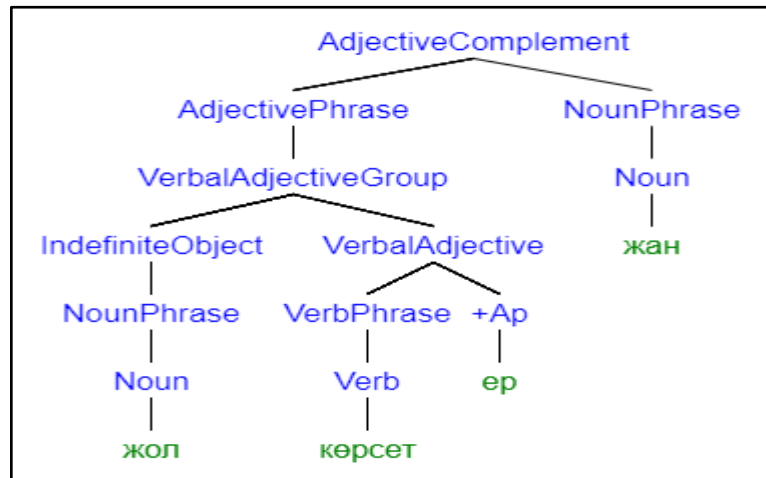
Şekil 3.8: Sıfat tamlaması.

3.3.3.3. Sıfat-Fiil Grubu (Verbal Adjective Group)

Bir sıfat-fiil ve tamlayıcı(lar)dan oluşan kelime grubudur. Bu grup içerisinde sıfat-fiil genellikle sonda bulunur. Yüklem görevi yapan sıfat fiilin anlamı, özne, nesne veya zarf unsurlarıyla tamamlanır. Şekil 3.9’de sıfat-fiil grubuna örnek verilmiştir.

VJG → VJ
 VJG → IOBJ VJ
 VJG → DOBJ VJ

жол	көрсетер	жан
жол	көрсет+Ap	жан
yol	gösterecek	kimse



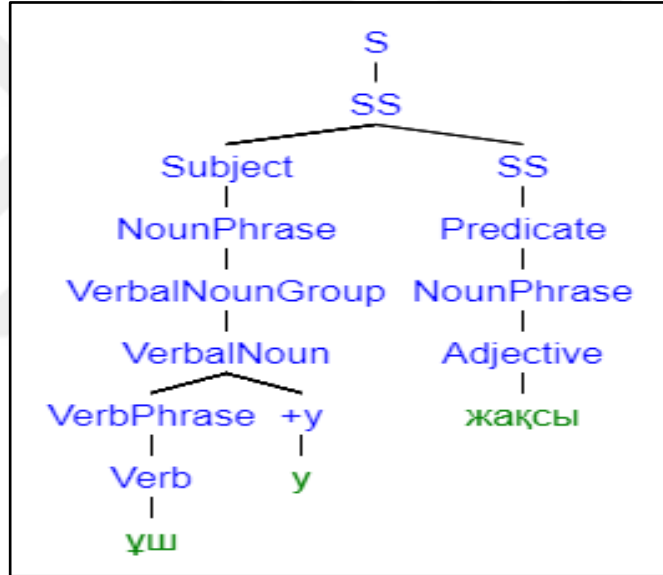
Şekil 3.9: Sıfat-fiil grubu

3.3.3.4. İsim-Fiil Grubu (Verbal Noun Group)

İsim fiil grubu (VNG), bir isim fiilden ve ona bağlı bir veya daha fazla tümleçten oluşur. Bir fiil +y ekini aldığımda isim fiil olur. Tamamlayıcılar (bileşenler) özne, nesne, dolaylı nesne veya zarftır. Şekil 3.10’de isim-fiil grubuna örnek verilmiştir.

VNG → VN
 VNG → IOBJ VN
 VNG → DOBJ VN

ўшу жақсы
 ўш+y жақсы
 uçmak güzel(dir)



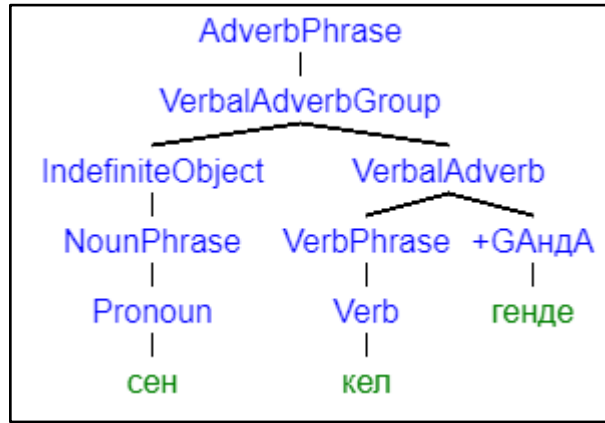
Şekil 3.10: İsim-fiil grubu.

3.3.3.5. Zarf-Fiil Grubu (Verbal Adverb Group)

Fiillerin hareket halini ifade eden fiil şekilleridir ve Şekil 3.11’de bunun bir örneği verilmiştir.

VAG → VA
 VAG → IOBJ VA
 VAG → DOBJ VA

сен келгенде
 сен кел+GANDA
 sen geldiğinde



Şekil 3.11: Zarf-fiil grubu.

3.3.3.6. Tekrar Grubu (Reduplication Group)

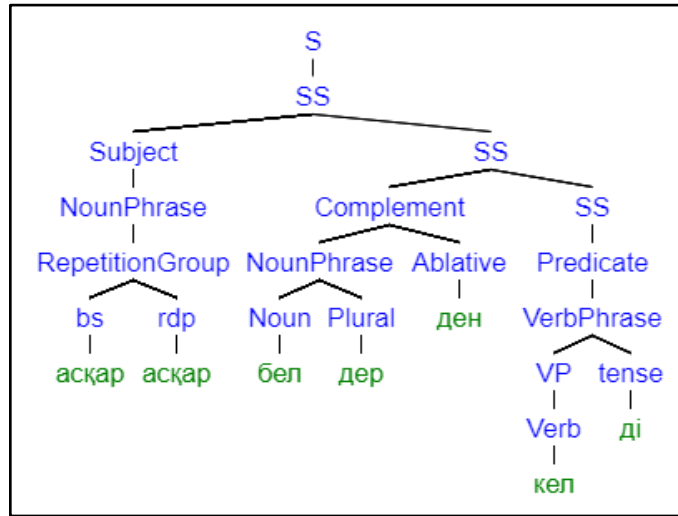
Bir nesneyi, bir hareketi karşılamak için eş görevli kelimelerden oluşan kelime grubudur. Kelimeler arasında şekil ve anlam ilişkisi vardır. Unsurları aynı olan, yakın olan ve zıt olan tekrar grupları olmak üzere 3 tip bulunur. Şekil 3.12’de bu tiplerden unsurları aynı olan tekrar grubuna örnek verilmiştir.

RG → base redup

асқар-асқар
асқар-асқар
yüksek yüksek

белдерден
бел+дер+ден
tepelerden

келді
кел+ді
geldi



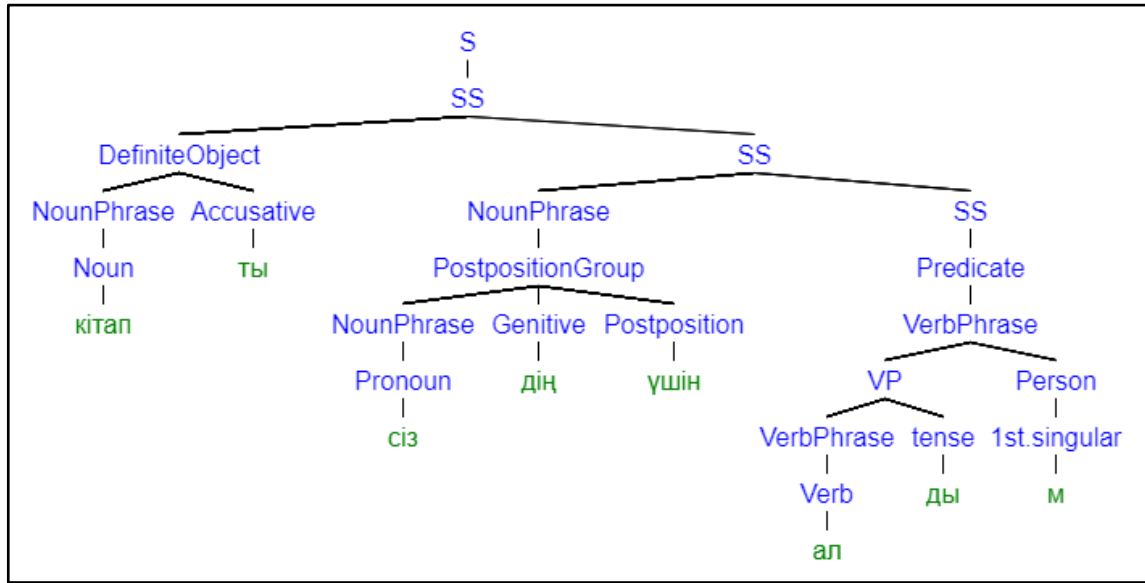
Şekil 3.12: Tekrar grubu.

3.3.3.7. Edat Grubu (Postposition Group)

Bir isim unsuru ile bir çekim edatının oluşturduğu kelime grubudur ve Şekil 3.13’da buna örnek verilmiştir.

- PG → NP posp
- PG → NP dat posp
- PG → NP gen posp
- PG → NP abl posp

кітапты	сіздің	үшін	алдым
кітап+ты	сіз+GEN	үшін	ал+TENSE+PERSON (1sg)
kitabı	sizin	için	aldım



Şekil 3.13: Edat grubu.

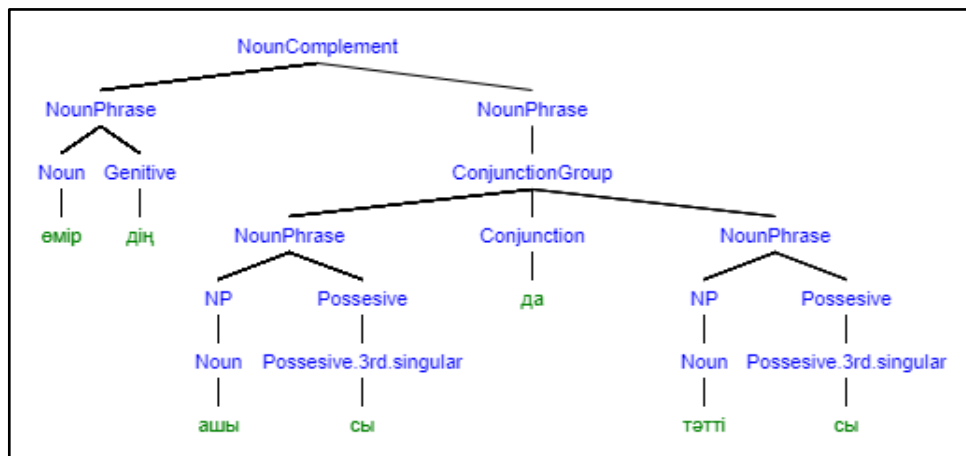
3.3.3.8. Bağlama Grubu (Conjunction Group)

Bağlaçlarla birbirine bağlanmış iki ya da daha fazla isim unsurundan oluşan kelime grubudur.

Bu kelime grubunun bir örneği Şekil 3.14’de verilmiştir.

CG → NP conj NP
 CG → NP cm CG
 CG → NP ins posp

өмірдің	ашысы	да	тәттісі
өмір+GEN	ашы+POSS (p3sg)	да	тәтті+POSS (p3sg)
hayatın	acısı	ve	sevinci



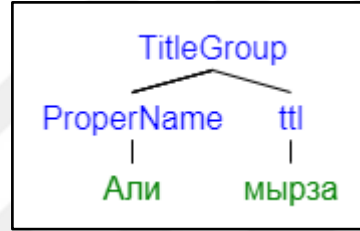
Şekil 3.14: Bağlama grubu.

3.3.3.9. Unvan Grubu (Title Group)

Bir şahıs ismi ve bir unvan ya da akrabalık isminden oluşan gruptur. Şekil 3.15’de isim ve unvandan oluşan bir örnek verilmiştir.

TG → prop ttl
 TG → ttl prop
 TG → PNG ttl
 TG → ttl PNG

Али мырза
 Али мырза
 Ali bey

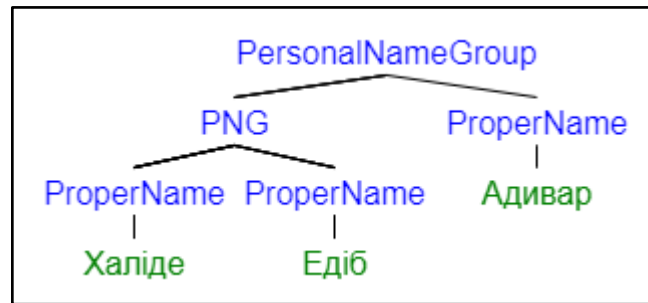


Şekil 3.15: Unvan grubu.

3.3.3.10. Özel İsim Grubu (Person Name Group)

İki ya da daha fazla özel isimden oluşan kelime grubudur. Aşağıda Halide Edib Adıvar örneği Şekil 3.16’da verilmiştir.

PNG → prop prop
 PNG → PNG prop

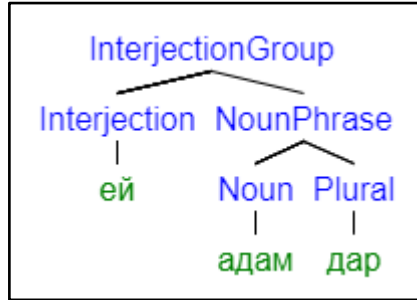


Şekil 3.16: Özel isim grubu.

3.3.3.11. Ünlem Grubu (Interjection Group)

Bir ünlem ve bir veya birden fazla isimden oluşan kelime grubudur. Bu kelime grubuna örnek Şekil 3.17’de verilmiştir.

IG	→	intj NP
IG	→	IG cm NP
ей		адамдар
ей		адам+PLU
ey		insanlar

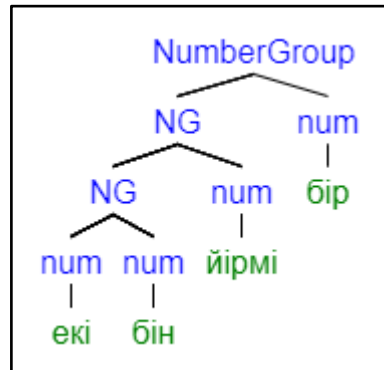


Şekil 3.17: Ünlem grubu.

3.3.3.12. Sayı Grubu (Number Group)

Basamak sistemine göre sıralanmış sayı isimleri topluluğudur. Sayı grubu cümle içerisinde isim ve sıfat görevi görürler. Sayı grubuna örnek Şekil 3.18’de verilmiştir.

NG	→	num num
NG	→	NG num



Şekil 3.18: Sayı grubu.

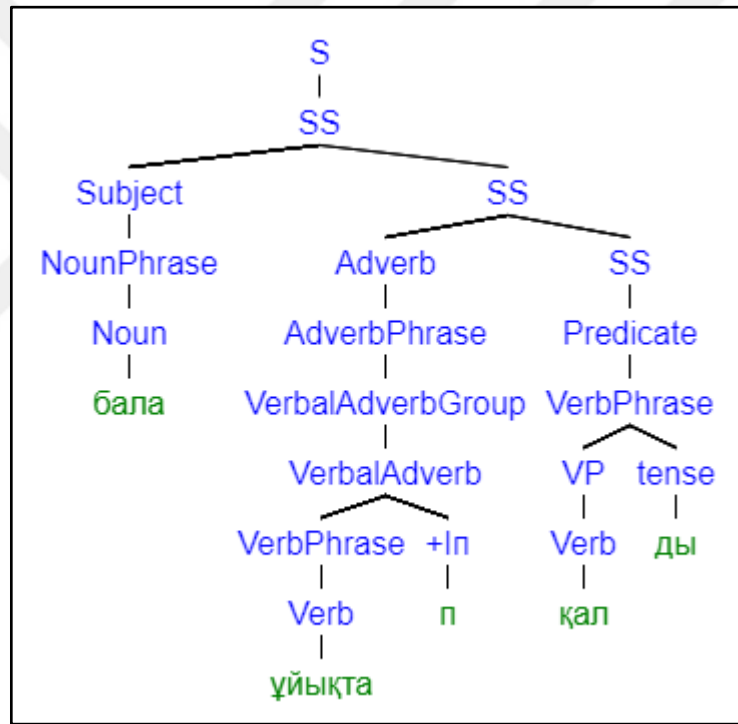
3.3.3.13. Birleşik Fiil (Compound Verb Group)

Bir hareketi karşılamak veya hareketi tasvir etmek için yan yana gelmiş kelimeler topluluğudur. Birleşik fiillerin isim unsuru bir sıfat fiil olabilir. bol- ‘olmak’ yardımcı fiili sıfat fiillerle birlikte kullanılmaktadır ve niyet ifade eden bir eylemi karşılar. Bir hareketi tasvir etmek için yan yana gelmiş birleşik fiiller ise zarf eki taşıyan bir fiille tasvir edilen fiille birleşmesinden oluşur.

CVG → NP XVP
 XVP → auxv
 XVP → XVP psv
 XVP → XVP neg
 XVP → XVP tense
 XVP → XVP Prs
 XVP → XVP aux
 XVP → XVP mood

şeklinde çoğaltılabilir. Şekil 3.19’de buna örnek verilmiştir.

бала	ұйықтап	қалды
бала	ұйықта+Іп	қал+TENSE
çocuk	uyuya-	kaldı



Şekil 3.19: Birleşik fiil grubu.

4. BULGULAR

4.1. KAZAKÇA BİÇİMBİLİMSEL ANALİZİNİN NÜVE'DE GERÇEKLENMESİ

Tezin bu aşamasında Kazakça dilinin morfolojik analizi yapılmıştır. Morfolojik analiz, bir dilde verilen bir kelimenin kök ve eklerine ayrılarak özelliklerinin incelenmesidir. Morfolojik analiz Türkçe, Kazakça, Kırgızca gibi sondan eklemeli diller için önemli bir yere sahiptir. Ayrıca cümle analizi, anlam analizi ve makine çevirisi gibi doğal dil işlemenin en önemli çalışmaları açısından önemli bir yere sahiptir ve bu çalışmaların temel adımını oluşturur.

Bir dilin morfolojik analizinin yapılırken dilin ortografik kurallara, morfotaktik kurallara, sözlük ve bir morfolojik çözümleyiciye ihtiyaç duyulmaktadır. Literatüre bakıldığında Türkçe ya da Türki diller için yapılmış morfolojik çözümleyicilerin başında Zemberek gelmektedir.

Bu tezde Kazakça dili için açık kaynaklı morfolojik çözümleyici olan Nüve kullanılmıştır. Nüve sondan eklemeli diller için tasarlanan Yukarıdan-Aşağıya (top-down) biçimbilimsel üretici ve çözümleyicidir. Ayrıca Türkçenin morfolojisi içerisinde hazır halde bulunmaktadır. Ayrıca Nüve kullanılarak Türki dillerden Kırgızca (Yiner, 2016) ve Kırım Tatarcası'nın (Sanli, 2018) morfolojik analizi yapılmıştır. Nüve çatısı kullanılarak Nüve geliştirilmiştir. Nüve web tabanlı bir morfolojik çözümleyici ve makine çeviri sistemi olarak kullanıma açılmıştır⁴.

4.1.1. Nüve Biçimbilimsel Çözümleyici

Nüve'de bir dil tanımlanırken 4 dosyaya ihtiyaç duyulmaktadır:

1. Ekler (Suffixes) sözlüğü
2. Kökler (Roots) sözlüğü
3. Ortografik kurallar XML dosyası
4. Morfotaktik kurallar XML dosyası

Kazakça dilinde bulunan tüm ekler *ekler* dosyasında tanımlanmıştır. Her ekin *id* si, *adı*, *türü*, *lexical formu*, varsa *grubu* bu ekin tüm *yüzeyleri* ve bu yüzeylerin dönüşümleri için gerekli olan *ortografik kurallar* ve özel durumlar için kullanılan *flag* özellikleri bulunmaktadır. Tablo 4.1'de Kazakça Ekler dosyasının bir bölümü verilmiştir.

⁴ <http://www.nuvestudio.com/>

Tablo 4.1: Ekler sözlüğü.

Id	Lexical Form	type	Grup	flags	rules	surfaces
NI_PLURAL_LAR	LAp	I			Transformation_A, Transformation_L	лар,лер,дар, дер,тер,тар
NI_CASE_GEN_NIN	NI _H	I	NOUN_CASE		Transformation_N, Transformation_I	ның,нің,ды ң,дің, тың,тің
NI_CASE_ACC_NI	NI	I	NOUN_CASE		Transformation_N, Transformation_I	ны,ні,ды,ді, ты,ті
VI_TENSE_PAST_DI	DI	I	VERB_TENSE		Transformation_D, Transformation_I	ды,ді,ты,ті
VI_TENSE_CONT_A	A	I	VERB_TENSE		Transformation_I, Special_A	а,е,й

Kökler (roots) sözlüğünde ise dildeki tüm kökler (gövdeler, kelimeler) bulunmaktadır. Her kök için **root**, metin içerisindeki tüm hallerini içeren **surfaces** özelliği, türünü (POS) belirleyen **id** si, özel durumlar için **flag** özelliği, ortografik kuralları içeren **rules** özelliklerinden oluşur. Bu özelliklerin olması bir dilin tanımlanması için yeterlidir. Fakat uygulama sırasında farklı amaçlar için bunların dışında da her girdi için başka özellikleri belirten sütunlar da eklenmiştir. Tablo 4.2’de *Kökler* sözlük dosyasından bir kesit verilmiştir.

Tablo 4.2: Kökler sözlüğü.

root	surfaces	Id	flags	rules
абажур		NOUN	noun	
адам		NOUN	noun	
ауыз	ауз	NOUN	noun	LAST_VOWEL_DROP
басқыншылық		ADJECTIVE	adjective	
баста		VERB	verb	
абстракттік	абстракттіг	NOUN	noun	MUTATION_kg
кітап	кітаб	NOUN	noun	MUTATION_pb
мен	ма,мені,ме	PERS_PRONOUN	pers_pronoun, PP1	SPECIAL_PPNOUN
ол	о,оны	PERS_PRONOUN	pers_pronoun, PP3	SPECIAL_OL
пайдашылдық	пайдашылдығ	NOUN	noun	MUTATION_qg

Bir ekin bir gövdeye eklendiğinde, ek üzerinde ya da eklendiği gövdede bazı değişimler olabilir. Bu değişimi sağlayan kurallar ortografik kurallardır. Dilin tanımlanmasında gereken üçüncü dosya ise ortografik kurallarını tanımlandığı dosyadır. Bu aşamada Kazakça dilinin ortografik kuralları iki düzeyli olarak tanımlanmıştır. Ortografik kuralların tanımlanmasını kolaylaştırmak için belirli harf kümelerini kullanılmıştır. Bu harf kümelerini gruplamak için

büyük harfler olarak gösterdiğimiz meta-harfler kullanılmaktadır. Aşağıda D meta harfinin (d ve t harflerini temsil eder) iki düzeyli kural olarak tanımlanması verilmiştir.

$$\begin{aligned} D:\text{д} &\Leftrightarrow [V|C_t] + __ \\ D:\text{т} &\Leftrightarrow C_{ts} + __ \end{aligned}$$

Bu kurala göre D meta harfi eklendiğin gövdenin son harfine göre *d* veya *t* harfine dönüşmektedir. Eğer gövde sert sessiz harfle (C_{ts}) bitiyorsa т (t) ye, ünlü (V) veya sert olmayan ünsüzle C_{ts} bitiyorsa д (d) ye dönüşmektedir. Aşağıda bu dönüşümlere örnek verilmiştir.

көк+DA	көк+DA	N(gök)+NNI_LOC
көк0те	көк0те	gökte
көкте	көкте	
қала+DAn	qala+DAn	N(şehir)+NNI_ABL
қала0дан	qala0дан	şehirден
қаладан	qaladan	

Nüve’de ortografik kurallar XML formatında tanımlanmaktadır. Nüve’de D meta harfinin değişimi gösteren ortografik kuralın tanımı Şekil 4.1’de verilmiştir. Her kural, birbirinden farklı **rule id** ve kuralın açıklamasının verildiği **description** taglarından oluşmaktadır. Bu ortografik kuralın hangi şart(lar) altında gerçekleştiğini belirleyen **conditions** tagi bulunmaktadır. Ortografik kuralın uygulanması sonucunda oluşan değişim **transformation** tagi arasında yazılır. Bu kural gibi dildeki diğer ortografik kurallar da XML formatında yazılarak Nüve’de tanımlanmıştır. Bazı ortografik kuralların diğer kurallardan önce uygulanması gerekir. Bunun için her ortografik kural için uygulandığı seviyeyi (toplamda 5 seviyeden oluşur.) göstermek için **phase** özelliği eklenmiştir. Örneğin lexical düzeyde verilen kök ve eklerin birleştirilmesi 1. seviyede, meta harflerinin dönüşümleri 2. seviyede, bazı ünsüz yumuşama kuralları 3. seviyede, bazı özel karakterlerin eklenmesi veya silinmesi (kesme işareti ekleme/çıkarma, alt tire simgesinin silinmesi vb.) ise 5. seviyede gerçekleşmektedir. Verilen örneğe göre, D meta harfinin değişmesi kuralı morfolojik üretimin 2. seviyesinde gerçekleşmektedir.

```

<rule id="TRANSFORMATION_D" phase="2">
  <description>This rule gives info about meta-letter D changes (d or t)</description>
  <transformation morpheme="This" action="Replace" operandOne="D" operandTwo="т" flag="all">
    <conditions flag="Or">
      <condition morpheme="Previous" operator="LastLetterEquals" operand="ккпстфхчшщ" />
    </conditions>
  </transformation>
  <transformation morpheme="This" action="Replace" operandOne="D" operandTwo="д" flag="all" />
</rule>

```

Şekil 4.1: Ortografik kurallar XML dosyası.

Dilin tanımlanması için gerekli olan bir diğer dosya ise, o dildeki isim/fiil köklerine eklerin gelme sırasını belirleyen morfotaktik kurallar dosyasıdır. Türkçede olduğu gibi Kazakça dilinde de ekler bir sözcük içerisinde soldan sağa doğru belirli bir sıraya göre eklenmektedir. Örneğin, çoğul ekinin iyelik ekinden önce isim köküne eklenmesi, hal eklerinin ise direk isim köküne veya bu eklerden sonra ekleniyor olması gibi durumlar verilebilir. Bu sıralanma Sonlu Durum Makinaları ile gösterilmektedir.

Şekil 3.1’de isim gövdeleri için çizilmiş olan Sonlu Durum diyagramından bir kesit verilmiştir. Şekil 3.3 ve Şekil 3.4 ise fiil kök veya gövdeleri için çizilmiştir. İsim ve fiil gövdeleri için çizilen bu Sonlu Durum Makinaları birbirine bağlıdır, daha iyi anlaşılması açısından bölünerek verilmiştir.

```

<suffixGroups>
  <suffixGroup name="NOUN_POSSESSIVE">
    <suffix>NI_POSSESSIVE_1sg_(I)m</suffix>
    <suffix>NI_POSSESSIVE_2sg_(I)ñ</suffix>
    <suffix>NI_POSSESSIVE_3sg_(s)I</suffix>
    ...
  </suffixGroup>
  <suffixGroup name="NOUN_CASE">
    <suffix>NI_CASE_GEN_NIn</suffix>
    <suffix>NI_CASE_ACC_NI</suffix>
    <suffix>NI_CASE_LOC_DA</suffix>
    ...
  </suffixGroup>
  <suffixGroup name="VERB_TENSE">
    <suffix>VI_TENSE_PAST_DI</suffix>
    <suffix>VI_TENSE_CONT_A</suffix>
    <suffix>VI_TENSE_FUT_MAK</suffix>
    ...
  </suffixGroup>
  ...
</suffixGroups>
<graph>
  <source id="NOUN">
    <targetGroup id="NOUN_POSSESSIVE" />
    <targetGroup id="NOUN_CASE" />
    <targetGroup id="ND_NOUN" />
    ...
  </source>
  <source id="VERB">
    <targetGroup id="VERB_TENSE" />
    <targetGroup id="VERB_MOOD" />
    <targetGroup id="VERBAL_ADVERB" />
    ...
  </source>
  ...
</graph>

```

Şekil 4.2: Morfotaktik XML dosyası.

Nüve’de morfotaktik kurallar XML formatında tanımlanmaktadır. Öncelikle uzun uzun tanımlamaları minimize etmek için ek grupları oluşturulur (Bunların daha önce ekler sözlüğünde de grupları belirlenmiştir.). **suffixGroup** diye adlandırılan her bir ek grubuna dâhil olan tüm ekler **suffix** etiketleriyle eklenir. Daha sonra her kök/ gövde türlerine mantıksal bir kelime oluşturmak için eklenen tüm ekler **target** olarak tanımlanır. Bir sonraki aşama ise artık morfotaktik grafının şekillendirilmesidir. Her bir ek ve root türleri graf için düğümdür ve **source** etiketiyle tanımlanır. Mevcut **sourcedan** diğer **source** a geçiş, yapım veya çekim ekleriyle sağlanır. Şekil 4.2’da Nüve’de tanımlanan Kazakça morfotaktik XML dosyasından bir kesit verilmiştir.

4.1.2. Nüve’de Kazakça Morfolojik Analiz Örnekleri

Universal Dependency⁵ veri tabanında Kazakça için hazırlanmış test ve train veri setleri (Makazhanov ve diğ., 2015, Tyers ve Washington, 2015) kullanılarak Nüve’de çözümlenerek, veri setinden her kelime için olası morfolojik çözümleri bulunmuştur. Bu veri setinde 1073 cümle, 10000 in üzerinde kelime (token) vardır. Kazakça ekler bakımından zengin olduğu için bulunan morfolojik çözümlemelerin sayısı da buna bağlı olarak artmıştır. Bir kelime için en fazla 5 morfolojik çözüm dikkate alınmıştır. Ayrıca doğru olan morfolojik çözüm ilk morfolojik çözüm olacak şekilde veri setinde işaretlenmiştir. Çözümleme yapılırken yaklaşık olarak 24000 girdiden oluşan Kazakça kökler sözlüğü (roots) ve 145 girdiden oluşan ekler sözlüğü (suffixes) oluşturulmuştur. Kullanılan iki düzeyli ortografik kurallar ve morfotaktik SDM bir önceki bölümde verilmiştir.

Kazakçanın morfolojik analizi masaüstü uygulaması olan Nüve’de derlenmiştir. Ayrıca Nuve geliştirilerek web tabanlı morfolojik çözümleyici⁶ olarak kullanılma açılmıştır. Kazakçanın dil dosyaları web uygulamasına taşınmıştır. Böylece web ortamında Kazakça morfolojik çözümleme ve üretme kullanıcıların kullanımına açılmıştır.

Aşağıda Şekil 4.3’de verilen cümlenin her bir birimi için (oluşturan her kelime) morfolojik çözümleri verilmiştir. “мал орнына сатылып, жаны сүймеген жанға жар болатын қазақтың сансыз қыздарының ішінен әлдеқалай өзінің сүйгеніне тиген шолпан шынында бастапқы жылы бала керек қылмады.” cümlesi için her kelimenin morfolojik çözümü hemen alt satırında verilmiştir:

⁵ https://universaldependencies.org/treebanks/kk_ktb/index.html

⁶ <http://www.nuvestudio.com/>

мал	мал+NOUN
орнына	орын+NOUN+NI_POSSESSIVE_3sg_(s) I+NI_CASE_DAT_GA
сатылып	сат+VERB+VD_PASS_I1+VERBAL_ADVERB_Ip
,	,+PUNCTUATION
жаны	жан+NOUN+NI_POSSESSIVE_3sg_(s) I
сүймеген	сүй+VERB+FD_NEGATIVE_mA+VI_TENSE_PAST_GAn
жанға	жан+NOUN+NI_CASE_DAT_GA
жар	жар+NOUN
болатын	бол+VERB+VI_TENSE_PAST_AtIn
қазақтың	қазақ+NOUN+NI_CASE_GEN_NIn
сансыз	сансыз+ADJECTIVE
қыздарының	қыз+NOUN+NI_PLURAL_LAr+NI_POSSESSIVE_3sg_(s) I+NI_CASE_GEN_NIn
ішінен	іш+NOUN+NI_POSSESSIVE_3sg_(s) I+NI_CASE_ABL_DAn
әлдеқалай	әлдеқалай+PRONOUN
өзінің	өз+PRONOUN+NI_POSSESSIVE_3sg_(s) I+NI_CASE_GEN_NIn
сүйгеніне	сүй+VERB+VERBAL_ADJECTIVE_GAn+NI_POSSESSIVE_3sg_(s) I+NI_CASE_DAT_GA
тиген	ти+VERB+VERBAL_ADJECTIVE_GAn
шолпан	шолпан+NOUN
шынында	шынында+ADVERB
бастапқы	бастапқы+ADJECTIVE
жылы	жыл+NOUN+NI_POSSESSIVE_3sg_(s) I
бала	бала+NOUN
керек	керек+MODAL
қылмады	қыл+VERB+FD_NEGATIVE_mA+VI_TENSE_PAST_DI
.	.+PUNCTUATION

Şekil 4.3: Morfolojik çözüm örneği.

4.1.3. Kazakça Biçimbilimsel Analiz Sırasında Bulunan Özel Durumlar ve İstisnalar

Bu bölümde Kazakça morfolojik analiz yapılırken karşılaşılan problemler ve varsa çözümleri hakkında bilgi verilmiştir.

Kazakça bazı ekler, eklendiği kelimedenden ayrı olarak yazılmaktadır. Örneğin Türkçe soru ekinde +mI (+mı, +mi, +mu, +mü) olduğu gibi Kazakça soru eki de +MA (+ma, +me, +ba, +be, +pa, +pe) eklendiği gövdeden ayrı yazılmaktadır. Bir morfolojik çözümleyicide arada boşluk olması durumunda bu kelimeleri birbirinden ayrı iki kelime olarak düşünmektedir. Nüve’de ise bu problemin çözümü olarak özel bir kural yazılmıştır. Ekler sözlüğünde bu tarz eklerin başına ‘_’ (alt tire) simgesi konulmuştur. Morfolojik üretimin (morphological generation) son seviyesinde bu simge silinir. Yüzey formda soru eki veya bunun gibi diğer ekler kelimedenden ayrı yazılmış olarak görülür.

Kazakçada bulunan olumsuzluk eki +MA eki ile fiilden isim yapan yapım eki olan +MA ekleri birbirinin aynısı gibi durmaktadır. İkisi de fiil kök veya gövdelerine eklenir. Örneğin аспа (aspa, asma) kelimesinin olumsuzluk ekiyle yapılan emir kip çekimi veya fiilden isim olarak türetilen bir isim morfolojik çözümü olabilir. Bu iki çözümün hangisinin seçileceğine verilen cümleye bakılarak karar verilir.

Bazı eklerin yapısal (lexical) formları farklı olup yüzey (surface) formlarından bazıları aynıdır. Örneğin isimden sıfat yapan yapım eki +LI (+лы, +лі, +ды, +ді, +ты, +ті) ekiyle belirtme hal eki +NI (+ны, +ні, +ды, +ді, +ты, +ті) eklerinden yüzeydeki formlarına bakıldığında ortak formlar görülebilir. Morfolojik çözümleyicimiz iki çözümü de bulabilmektedir.

Kazakçada şart kipinde soru çekimi +MA soru eki haricinde +ше (+şe) ile de yapılmaktadır. Örneğin ‘yazsam mı’ anlamında ‘жазсам ше?’ veya ‘жазсам ба?’ kullanılabilir.

Kazakça şimdiki zaman eki olan +Ip (+іп) eki yanına тұр (tur), отыр (otur), жатыр (jätir) ve жүр (jür) kelimelerinden birini alarak çekimlenir. Fiilin zaman çekimi için gerekli olan şahıs ekleri ise bu yardımcı kelimelere eklenir. Ayrıca bu yardımcı kelimeler zaman ekinden ayrı olarak yazılır. Nüve’de bu zaman çekimini birbirinden ayrı yazıldığı için tek bir çekim olarak ele alınamadığı için zaman ekiyle bu yardımcı kelimeler alt tire simgesiyle birleştirilerek tek

bir ekmiş gibi ekler sözlüğüne eklenmiştir (+Іп_тұр, +Іп_отыр, +Іп_жатыр, +Іп_жүр şeklinde). Böylece bu zaman eki türünün 4 farklı eki varmış gibi varsayılmıştır.

Aynı durum gelecek zaman eki +ҒАЛІ (+GAlI) içinde yapılmıştır. Ek, тұр (tur), отыр (otur), жатыр (jatr) ve жүр (jür) yardımcı kelimeleriyle birlikte yazılmıştır. +ҒАЛІ ekiyle yapılan 4 farklı gelecek zaman eki var olduğu varsayıldı.

Kazakça istek kiplerinden birinde fiil +ҒІ (-ğı, -gi, -qı, -ki) eki, sonra iyelik eklerinden birini ve son olarak келеді (keledi) yardımcı kelimesini alarak çekimlenir. Bu istek kipinin, morfolojik üretim ve çözümleme kuralları eklenerek doğru şekilde çalışması sağlanmıştır.

Gereklilik kiplerinden birinde +УБ (+Uv) ekinden sonra iyelik eki ve yardımcı kelime керек (kerek) kullanılır. Bu kip çekimi de istek kipinin çekimindeki çözüm uygulanmıştır.

Romen rakamıyla yazılan kelimelerin morfolojik çözümü yapılamamaktadır. Morfolojik çözümünün olabilmesi için sözlük içerisinde geçmesi gerekmektedir. Fakat böyle bir durum yapılması mümkün değildir.

Özel isimlerin morfolojik analizinin yapılması için sözlüğe eklenmiş olması gerekir. Dilde geçen özel isimlerin sayısı çok olduğundan dolayı elle sözlüğe eklenmesi mümkün değildir. Bu daha sonra morfolojik çözümleyicinin farklı veri setlerinde test edilerek bu tarz kelimeler bulunduktan sonra eklenerek sorun kısmi olarak giderilebilir.

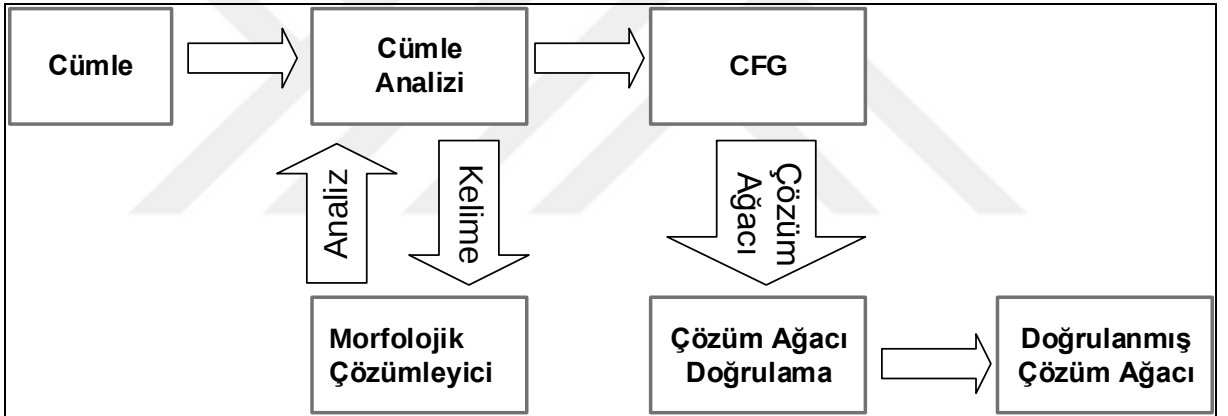
4.2. KAZAKÇA SÖZDİZİMSEL ANALİZİNİN NÜVE'DE GERÇEKLENMESİ

Bir dilin sözdizimsel analizinin yapılabilmesi için öncelikle o dilin gramer kurallarına ihtiyaç vardır. Kazakça gramer kurallarıyla sözdizimsel çözümleyiciyle Kazakça verilen bir cümle unsurlarına ayrıştırılarak, parse ağacı oluşturulabilir. Kazakça gramer kaynağı olarak Biray (2003), Biray ve diğ. (2015) ve Sevim (2003) kullanılmıştır. Sözdizimsel çözümleyici olarak Fatih parser (Zafer, 2011) referans alınarak Nüve sözdizimsel çözümleyici geliştirilmiştir. Nüve sözdizimsel çözümleyici içerisinde Türkçe, Türkmençe ve Özbekçe dilleri hazır halde bulunmaktadır.

4.2.1. Nüve Sözdizimsel Çözümleyici

Tezin ilk bölümünde morfolojik çözümleyici olarak Nüve kullanılmıştır. Nüve bir masaüstü uygulaması olup geliştirilerek web ortamına (<https://nuvestudio.com/>) taşınmıştır. Online Nüve’de Kazakça, Türkçe ve Özbekçe morfolojik analiz ve morfolojik üretim yapılabilmektedir.

Tezin bu bölümü olan, sözdizimsel analiz kısmında ise geliştirilen sözdizimsel çözümleyici Nüve kullanılmıştır. Nüve’de bir dilin sözdizimsel analizini yapabilmek için o dilin gramer kurallarına, veri setine (morfolojik olarak çözümlenmiş) ve çözümleme sonrasında bulunan çözümlerin doğruluğunu kontrolü için doğrulama (Parse tree validator) kurallarına, manuel tag (ekler) listesine ihtiyaç duyulmaktadır. Şekil 4.4’de sözdizimsel analiz için akış diyagramı verilmiştir.



Şekil 4.4: Sözdizimsel çözümleyici akış diyagramı.

Akış diyagramına göre, öncelikle cümlenin analizi yapılmaktadır. Bu analizden sonra bir morfolojik çözümleyici (Nüve) aracılığıyla parse edilen cümle/cümle parçacıklarının morfolojik çözümleri bulunur. Burada her kelime kök peşi sıra ekler dizisinden oluşmaktadır. Bu morfolojik çözümler kümesi, bir sözdizimsel çözümleyiciye verilir. Bu çözümleyici arka planda dilin gramer kurallarına bağlı olarak cümlenin sentaks analizini sağlayarak, o cümle için çözüm ağacı/ağaçları (parse tree(s)) elde eder. Bazı cümleler için ağaç sayısı çok fazla elde edilebilmektedir. Bunlardan bazılarını gramatik olarak doğru olsa da semantik açıdan doğru değildir. Bunu ayırt edebilmek için kurallara ihtiyaç vardır. Bu kurallar Parse Tree Validator aracıyla uygulanarak elde edilen çözüm ağaçları sayısını (gramer açısından doğru olarak)

minimize edilmektedir. Bu işlem sonucunda ilgili cümle için hem gramatik hem de semantik açıdan geçerli çözüm ağaçları elde edilir.

Tablo 4.3: Manuel tag ve eklerin eşleştirilmesi.

Ek	Tag	Ek	Tag
NOUN	noun	NI_POSSESSIVE_1sg_(I)m	p1s
PROPERNOUN	prop	NI_POSSESSIVE_2sg_(I)ñ	p2s
ADJECTIVE	adj	NI_POSSESSIVE_1pl_(I)mIz	p1p
TENSE	tense	PERS_POSS_1sg_m	1sg
ADVERB	adv	PERS_POSS_2sg_ñ	2sg
PRONOUN	pron	PERS_POSS_1pl_k	1pl
VERB	verb	VI_TENSE_PAST_DI	tense
PREPOSITION	posp	VI_TENSE_CONT_A	tense
NUMBER	num	VI_TENSE_FUT_MAK	tense
CONJUNCTION	conj	VI_TENSE_AORST_Ar	tense
QUES_MA	que	VI_MOOD_COND_sA	mood
NI_PLURAL_LAr	plu	VI_MOOD_COND_sA	mood
NI_CASE_GEN_NIn	gen	VI_MOOD_NECESS_Uv	mood
NI_CASE_DAT_GA	dat	VI_MOOD_IMPR_3SG_sIn	mood
NI_CASE_LOC_DA	loc	SECONDTENSE_COND_bolsa	aux
NI_CASE_ABL_DAn	abl	SECONDTENSE_NARR_edi	aux
NI_CASE_ACC_NI	acc	SECONDTENSE_PASTPRF_eken	aux
NI_CASE_ACC_N(3)	acc	VD_CAUS_DIr	+DIp
NI_CASE_INST_menen	ins	VERBAL_ADVERB_Ip	+In
NI_CASE_INST_men	ins	VERBAL_ADJECTIVE_GAn	+GAH
NI_REL_GI	rel	VERBAL_NOUN_U	+y
ND_NOUN_LIK	+LIK	FD_NEGATIVE_mA	neg
ND_ADJ_LI	+LI	ND_VERB_LA	+LA

Kazakçanın sentaks analizini için kullanılan veri seti, tezin morfolojik analiz kısmında kullanılmış olan veri setidir. Bu veri setinde her kelime, kök peşi sıra eklerin yazılması şeklinde hazırlanmıştır. Örneğin, *evim* kelimesi için morfolojik çözüm ‘ev/noun+NI_POSSESSIVE_1sg_(I)m’ şeklinde elde edilmiştir. Morfolojik çözümlemeyle elde edilen bu veri seti biraz değiştirilerek sözdizimsel çözümleyicinin kullanabileceği formata dönüştürülmüştür. Bu formatta, *evim* kelimesinin morfolojik çözüm çıktısı artık ‘ev/noun+im/p1s’ şeklindedir. Örneğin *KelimeKöküEk1Ek2* şeklinde bir kelime olduğunu varsayalım, *EkId* ise kelime içinde geçen ekin Nüve çözümleyicideki suffix dosyasındaki karşılığı, *Id* sidir. Nüve morfolojik çözümleyici çıktısı:

KelimeKökü+Tür+EkId1+EkId2

Sözdizimsel çözümleyici için gereken morfolojik çözüm formatı:

KelimeKökü/Tür+ Ek1/tag1+Ek2/tag2

şeklindedir. Dolayısıyla, Nüve morfolojik çözümleyicideki eklerle sözdizimsel çözümleyici manuel tagleri arasında eşleştirilme yapılmıştır. Bu eşleştirilmeden bir kesit Tablo 4.3’de verilmiştir.

Tag-ek eşleştirilmesinden sonra elde edilen düzenlenmiş veri setinde de her kelimenin olası tüm morfolojik çözümleri sekmeye ayrılarak yazılmıştır (örneğin тиген). Böylece sözdizimsel analiz için de veri seti oluşturulmuştur. Bu veri setinde geçen bir cümlenin morfolojik çözümü aşağıda Şekil 4.5’de verilmiştir.

Önceki veri setinde kelimenin doğru morfolojik çözümün ilk morfolojik çözümün doğru olacak şekilde düzenlenmiştir. Dolayısıyla yeni veri setinde de ilk çözümün doğru olduğu kabul edilmektedir.

мал	мал/noun
орнына	орын/noun+ы/p3s+на/dat
сатылып	сат/verb+ыл/+Іл+ып/+Іп
,	,+PUNCTUATION
жаны	жан/noun+ы/p3s
сүймеген	сүй/verb+ме/neg+ген/tense
жанға	жан/noun+ға/dat
жар	жар/noun
болатын	бол/verb+атын/tense
қазақтың	қазақ/noun+тың/gen
сансыз	сансыз/adj
қыздарының	қыз/noun+дары/p3p+ның/gen
ішінен	іш/noun+і/p3s+нен/abl
әлдеқалай	әлдеқалай/pron
өзінің	өз/pron+і/p3s+нің/gen
сүйгеніне	сүй/verb+ген/+GАн+і/p3s+не/dat
тиген	ти/verb+ген/+GАн ти/verb+ген/tense
шолпан	шолпан/noun
шынында	шынында/adv
бастапқы	бастапқы/adj
жылы	жыл/noun+ы/p3s
бала	бала/noun
керек	керек/MODAL
қылмады	қыл/verb+ма/neg+ды/tense
.	./PUNCTUATION

Şekil 4.5: Şekil 4.3’de verilen morfolojik çözümler örneğinin ek ve tag eşleşmesinden sonraki hali.

Sözdizimsel çözümleyici Nüve’de, gerekli olan bir diğer dosya ise dilin XML formatında yazıldığı gramer kurallarıdır (yani burada CFG kurallar, çözümleyicide derlenmesi için XML

formatında hazırlanmıştır.) .Gramer kurallarında, özne, yüklem, nesne, zarf, sıfat, tamlamalar (isim ve sıfat), sıfat-fiil, isim-fiil, zarf-fiil gibi cümleyi oluşturan her unsur için kurallar yazılmıştır. Her gramer kuralı *rule* tagleri içerisinde yazılmaktadır. Daha sonra kuralı (kategori) sağlayan kombinasyonlar *category* tagleri arasına yazılmaktadır. Örneğin Kazakça cümle içerisinde NP veya NP lerden oluşan bir kelime özne olabilir (Şekil 4.6).

```
<rule category="Sub">
  <category name="NP"/>
  <category terminal="true" name="cm"/>
  <category name="Sub"/>
</rule>
<rule category="Sub">
  <category name="NP"/>
</rule>
```

Şekil 4.6: Gramer.XML (Özne).

Kazakçada dolaylı nesne NPden oluşabilir. Ama dolaysız nesne ise ismin belirtme hal ekini almış olan NP den ve DObj lerden oluşabilir (Şekil 4.7).

```
<rule category="IObj">
  <category name="NP"/>
</rule>
<rule category="DObj">
  <category name="NP"/>
  <category terminal="true" name="acc"/>
</rule>
<rule category="DObj">
  <category name="NP"/>
  <category terminal="true" name="acc"/>
  <category terminal="true" name="cm"/>
  <category name="DObj"/>
</rule>
```

Şekil 4.7: Gramer.XML (Nesne).

Kazakçada yüklem (Şekil 4.8), VP veya soru ekini almış VP, NP veya soru ekini almış NP, var-yok-değil kelimelerinden biri (pre), ek almamış ya da şahıs eki almış NP olabilir. Bu ihtimaller çoğaltılabilir.

```

<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 terminal="true" name="pre"/> <!-- var, yok, değil -->
</rule>
<rule category="Pre">
  <category terminal="true" name="pre"/> <!--var mı, yok mu, değil mi -->
  <category name="Que"/>
</rule>
<rule category="Pre">
  <category name="NP"/><!--dir eki kazakçada olmadığı için direk isim yüklem olabilir-->
</rule>
<rule category="Pre">
  <category name="NP"/>
  <category name="Prs"/>
</rule>
<rule category="Pre">
  <category name="NP"/>
  <category terminal="true" name="aux"/>
  <category name="Prs"/>
</rule>

```

Şekil 4.8: Gramer.XML (Yüklem).

Kazakça NP (Şekil 4.9) ise isim, zamir, özel isim, çoğul ekini almış isim, isimden ya da sıfattan isim yapan yapım ekini almış isim, iyelik ekini almış isimler olabilir. Bu ihtimaller çoğaltılabilir.

Şekil 4.6, Şekil 4.7, Şekil 4.8 ve Şekil 4.9’de cümleyi oluşturan unsurlardan özne, nesne, yüklem ve isim unsuru için yazılmış gramer kurallarından bir bölüm verilmiştir. Bu kurallar cümle içerisinde geçen diğer unsurlar/kelime grupları için de ayrı ayrı yazılmıştır.

Nüve sözdizimsel çözümleyici, dilin gramer kurallarını kullanarak ilgili cümle için çözüm ağaçları (parse tree) oluşturur. Elde edilen bu çözüm ağaçlarının bazıları doğru çözüm ağacı olamayabilir. Bu yüzden bu çözüm ağaçlarının doğruluğunun test edilmesi gerekir. Çözümleyici bunu sağlamak için ayrı bir modül olan Parse Tree Validator modülü eklenmiştir.

Modülde yine gramer kuralları gibi, cümle yapısında olma(ma)sı gerekenler doğrulama (validation) kuralları olarak XML formatında yazılmıştır. Bu modül sonucunda doğrulanmış çözüm ağaçları elde edilir.

```
<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="noun"/>
  <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="+cI3"/>
</rule>
<rule category="NP">
  <category terminal="true" name="adj"/>
  <category terminal="true" name="plu"/>
</rule>
<rule category="NP">
  <category name="NP"/>
  <category name="Poss"/>
</rule>
<rule category="NP">
  <category name="NP"/>
  <category terminal="true" name="gen"/>
  <category terminal="true" name="rel"/>
</rule>
```

Şekil 4.9: Gramer.XML den bir bölüm (İsim unsuru-NP).

Şekil 4.10'de Validation.XML dosyasından bir kesit verilmiştir. Burada görüldüğü gibi *rule* taglari arasında kurallar yazılmıştır. İlk kural, Türkçede olduğu gibi Kazakçada da yüklem geçişli olup olmasına bağlı olarak cümlede nesne olur ya da olmaz. Yani eğer yüklem geçişsiz

fiillerden (nesne almayan fiil) birisiyse cümlede nesne olmaz. İkinci kuralda ise isim cümlelerinde (Predicate NP ise) nesne olmadığını belirtmek için yazılmıştır. Üçüncü kuralda ise, yine cümle içerisinde nesne olup olmamasıyla ilgili yazılmış bir kuraldır. Burada ise fiil (yüklem) edilgen olduğunda cümlede nesne yoktur. Örneğin “*temizlik yapıldı*” denildiğinde, burada *ne* sorusuna cevap olarak her ne kadar “*temizlik*” verilse de, “*temizlik*” bu cümlede özne görevindedir. Dördüncü kuralda ise özne-yüklem uyumluluğunu belirtmek için yazılmıştır. Özne hangi zamir ise, yüklem de o zamire uyumlu şahıs ekini almalıdır. Bu kuralda da yazıldığı gibi özne ben zamiri ise, yüklem 1.tekil şahıs ekini almalıdır.

```
<rule scope="SS" ignore="VJG, VNG, VAG">
  <description>Yüklem geçişsiz fiil ise cümlede nesne olmaz</description>
  <status type="cond" cat="Pre" op="hasNonTransitiveVerb"/>
  <status type="rest" cat="IObj" op="notExists"/>
  <status type="rest" cat="DObj" op="notExists"/>
</rule>
<rule scope="SS">
  <description>İsim cümlesinde nesne olmaz</description>
  <status type="cond" cat="Pre" op="startsWith" str="[Pre[NP]]"/>
  <status type="rest" cat="IObj" op="notExists"/>
  <status type="rest" cat="DObj" op="notExists"/>
</rule>
<rule scope="SS" ignore="VJG, VNG, VAG">
  <description>Yüklem edilgen ise nesne olmaz</description>
  <status type="cond" cat="Pre" op="includes" str="[psv[]]/>
  <status type="rest" cat="IObj" op="notExists"/>
  <status type="rest" cat="DObj" op="notExists"/>
</rule>
<rule scope="SS" ignore="VJG, VNG, VAG">
  <description>Özne BEN zamiri ise yüklem 1. tekil şahıs eki almalıdır</description>
  <status type="cond" cat="Sub" op="equals" str="[Sub[NP[pron[мен]]]]"/>
  <status type="rest" cat="Pre" op="includes" str="[Prs[1sg[]]]"/>
</rule>
```

Şekil 4.10: Validation.XML dosyasından bir kesit.

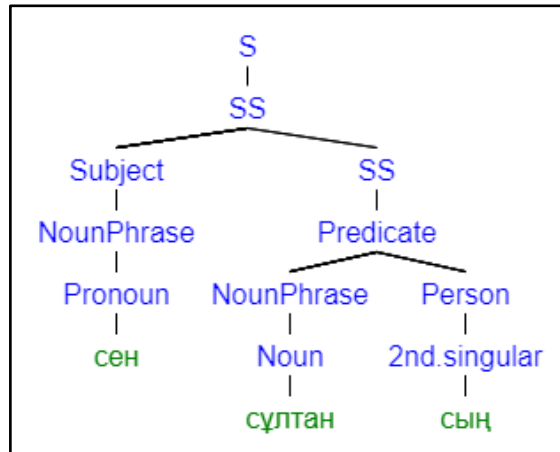
4.2.2. Nüve’de Kazakça Sözdizimsel Analiz Örnekleri

Bu bölümde Kazakça cümle türlerine göre (isim-fiil, olumlu-olumsuz-soru, basit-birleşik, kurallı-devrik) örnek cümleler ve bu cümlelerin çözüm ağaçları (parse tree) verilmiştir.

- **İsim cümlesi:** olumlu, olumsuz ve soru isim cümlesi örnekleri sırasıyla Şekil 4.11, Şekil 4.12 ve Şekil 4.13’de verilmiştir.

○ **Olumlu cümle:**

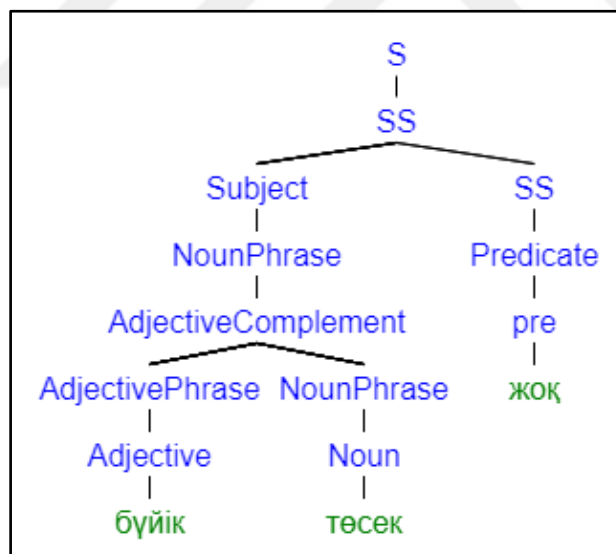
сен сұлтансың (sen sultansın)



Şekil 4.11: Olumlu isim cümlesi.

○ **Olumsuz cümle:**

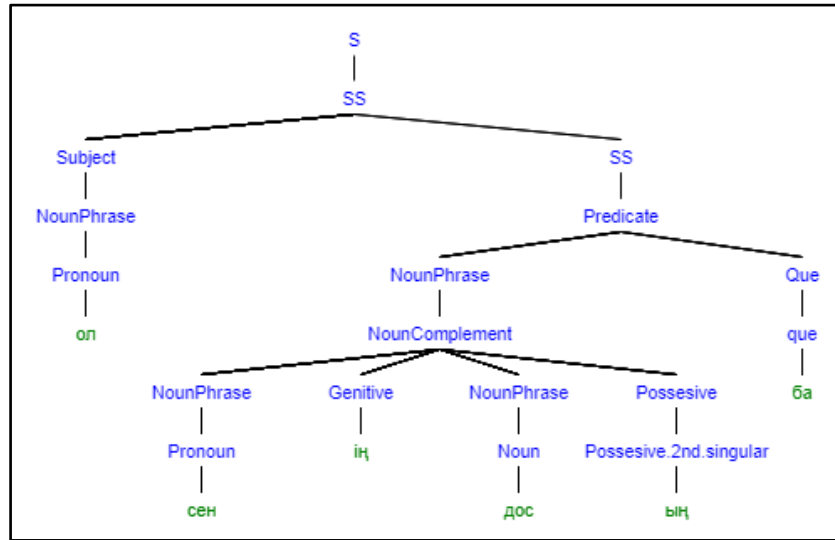
бүйік төсек жоқ (büyük yatak yok)



Şekil 4.12: Olumsuz isim cümlesi.

○ **Soru cümlesi:**

ол сенің досың ба? (o senin dostun mu?)

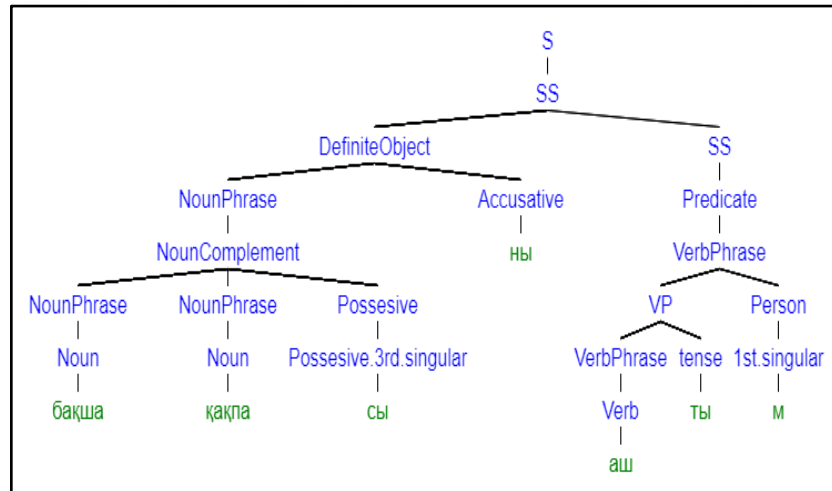


Şekil 4.13: Soru isim cümlesi.

- **Fiil cümlesi:** olumlu, olumsuz ve soru fiil cümlesi örnekleri sırasıyla Şekil 4.14, Şekil 4.15 ve Şekil 4.16’da verilmiştir.

- **Olumlu cümle:**

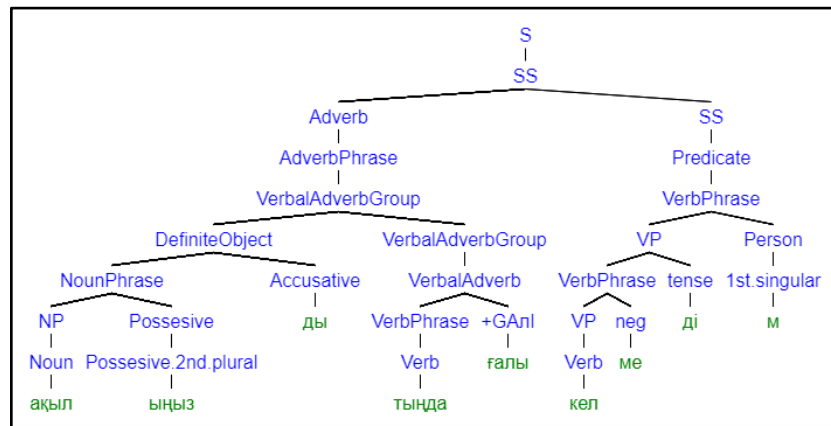
бақша қақпасыны аштым (bahçe kapısını açtım)



Şekil 4.14: Olumlu fiil cümlesi.

- **Olumsuz cümle:**

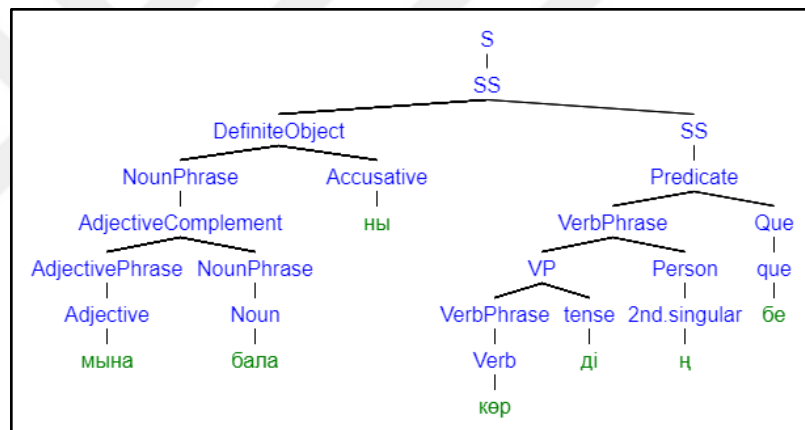
ақылыңызды тыңдағалы келмедім (nasihatınızı almaya gelmedim)



Şekil 4.15: Olumsuz fiil cümlesi.

○ **Soru cümlesi:**

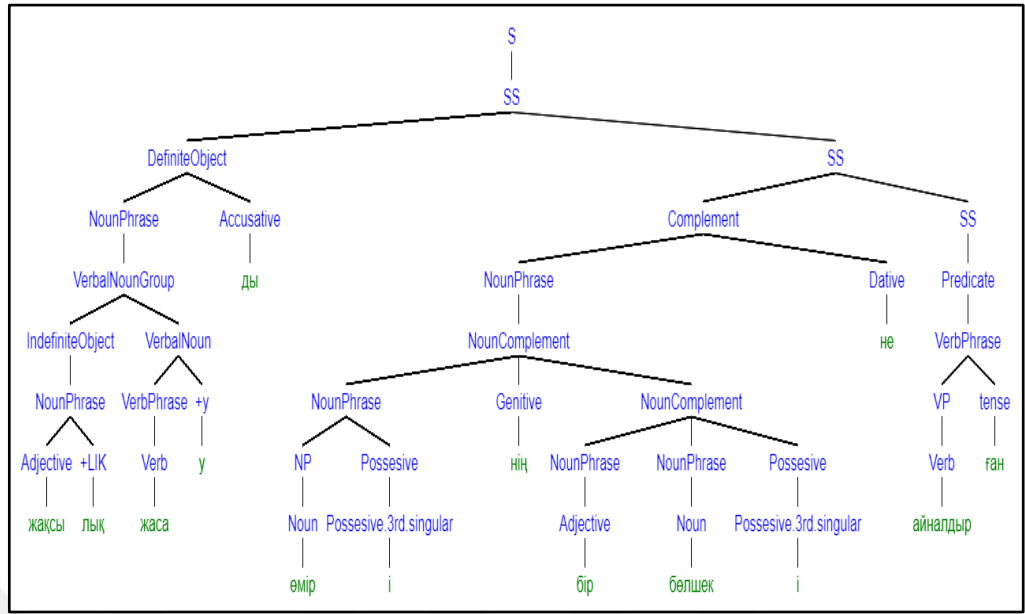
мына баланы көрдің бе? (Bu çocuğu gördün mü?)



Şekil 4.16: Soru fiil cümlesi.

- **Kurallı cümle:** Yüklemini sonda olan cümle türüdür ve Şekil 4.17’de kurallı cümle örneği verilmiştir.

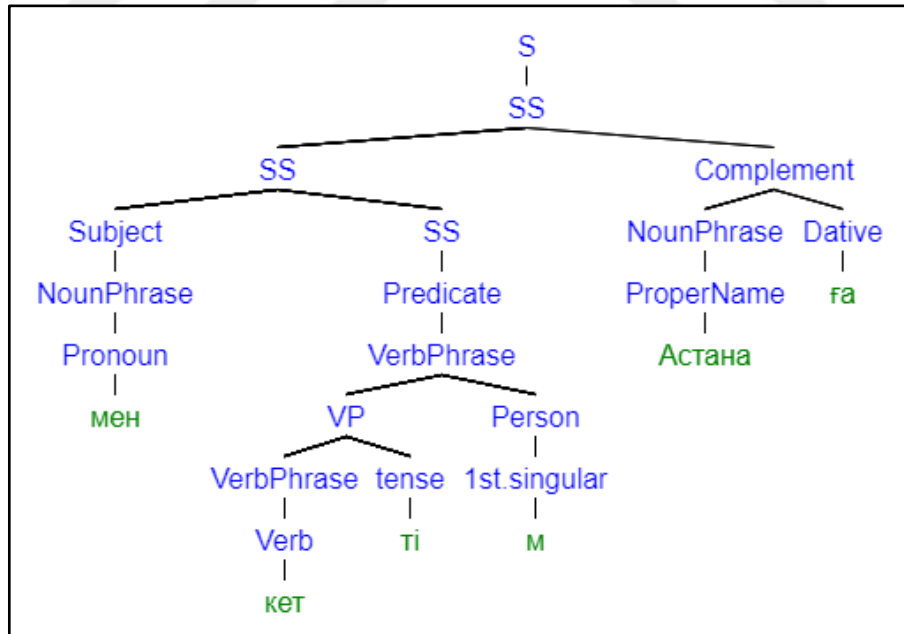
жақсылық жасауды өмірінің бір бөлшегіне айналдырған
(İyilik yapmayı hayatının bir parçası haline getirdi)



Şekil 4.17: Kurallı cümle.

- **Devrik cümle:** Yüklemi sonda olmayan cümle türüdür ve Şekil 4.18'de örneği verilmiştir.

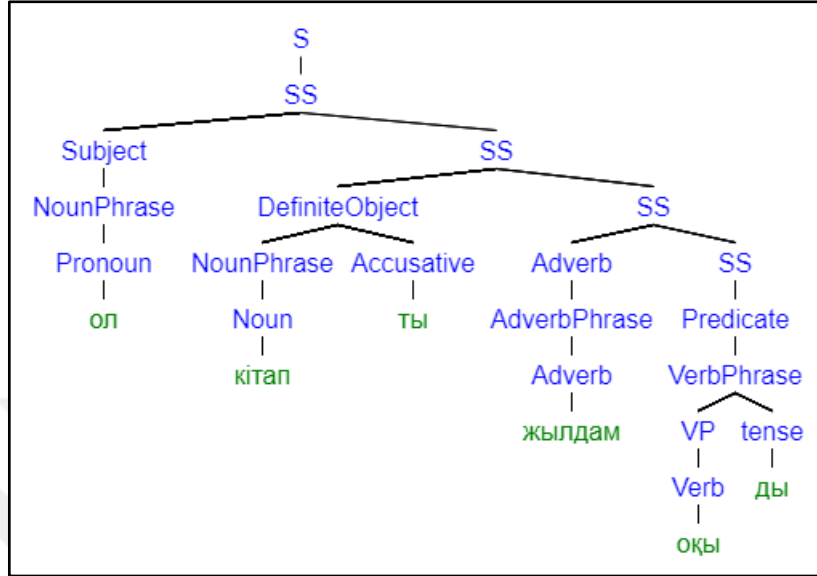
мен кеттім Астанаға (ben gittim Astana'ya)



Şekil 4.18: Devrik cümle.

- **Basit cümle:** Kazakça basit cümle örneği Şekil 4.19’da verilmiştir.

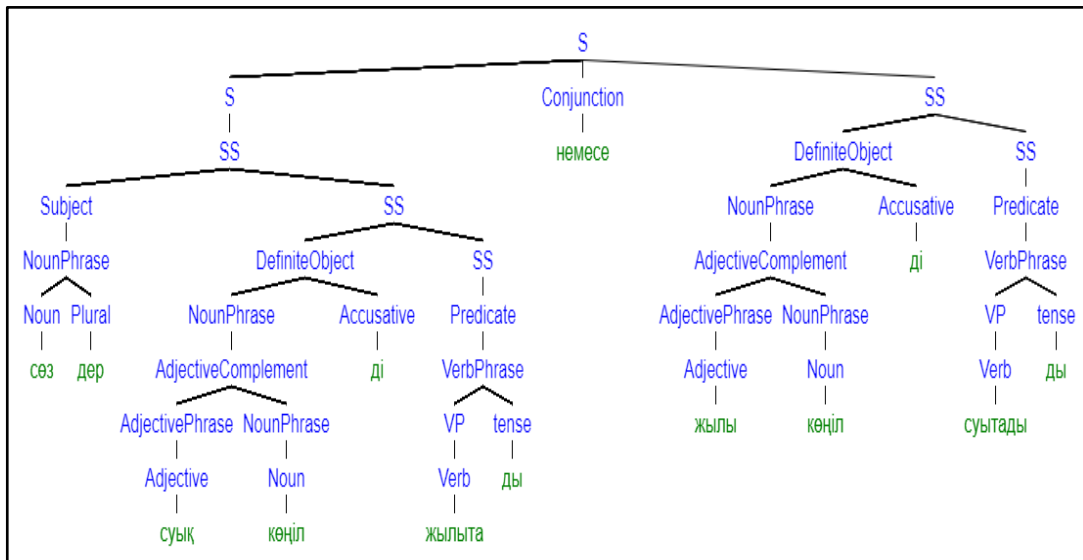
ОЛ КІТАПТЫ ЖЫЛДАМ ОҚЫДЫ (O kitabı hızlı okudu)



Şekil 4.19: Basit cümle.

- **Birleşik cümle:** Kazakça da bağlaçlarla oluşturulan birleşik cümle örneği Şekil 4.20’de verilmiştir.

сөздер суық көңілді жылытады **немесе** жылы көңілді суытады.
(sözler soğuk gönlü ısıtır veya sıcak gönlü soğutur.)



Şekil 4.20: Birleşik cümle.

4.2.3. Kazakça Sözdizimsel Analizi Sırasında Bulunan Özel Durumlar ve İstisnalar

Bu bölümde Kazakça sözdizimsel analiz yapılırken karşılaşılan problemler hakkında detaylı bilgi verilmiştir.

Kazakçada bazı ekler eklendiği gövdeden ayrı olarak yazılmaktadır. Örneğin bazı fiil çekimlerinde negatiflik anlamı emes/jok (değil/yok) kelimeleriyle sağlanırken bazıları ise +MA olumsuzluk ekiyle sağlanmaktadır. Emes/jok ile yapılan fiil çekimlerinde, bu yardımcı kelimeler önceki kelimedenden ayrı yazıldığı için sözdizimsel çözümleyici bunları ayrı iki parça olarak ele almaktadır. Ayrıca bazı zaman/kip çekimlerinde şahıs ekleri de bu yardımcı kelimelerden sonra gelmektedir.

Uygulama sırasında karşılaşılan problemlerden bir diğeri önceki bölümde +GI ekiyle yapılan istek kipinin çekimi verilebilir. Morfolojik çözümleme aşamasında ayrı yazılan yardımcı kelime (kerek) önceki fiil çekim grubu ile alınırken, sözdizimsel analiz aşamasında ise ayrı ele alınmıştır. Bunun gibi fiil çekiminde kullanılan yardımcı fiil/kelimeler morfolojik analizlerin bulunduğu dosyaya yardımcı kelimeler olarak kaydedilmiştir (kerek/aux şeklinde).

Kazakça bazı ekler eklendiği kelimeyle arasına – (tire) işareti gelmektedir. Örneğin Türkçe “55.” (elli beşinci) nin Kazakça karşılığı 55-ııı (55-şı) olarak metin içerisinde geçtiği görülmektedir. Bunun analizi (morfolojik dolayısıyla içerisinde geçtiği cümlelerin sözdizimsel analizi) yapılamamıştır.

Romen rakamıyla yazılan kelimelerin morfolojik çözümü yapılamamasından dolayı içerisinde geçtiği cümlelerin sözdizimsel analizi de yapılamamıştır.

Özel isimlerden dolayı bazı cümlelerin otomatik olarak sözdizimsel çözümlemesi yapılamamaktadır. Bu sorun, bu özel kelimeleri elle sözlüğe eklenmesiyle giderilmesi sağlanabilir. Fakat bu çözüm de uzun vadede mantıklı değildir.

5. TARTIŞMA VE SONUÇ

Bu tezde öncelikle Türki dillerden biri olan Kazakçanın gramer yapısı üzerinde araştırmalar yapılmıştır. Dil işlenirken en güncel gramer kaynakları referans alınmıştır. Bu kaynaklar sayesinde dilde bulunan ekler kapsamlı olarak ele alınmıştır. Bir dilin morfolojik analizinin yapılabilmesi için gerekli olan bir çözümleyici, gramer kuralları (yazım ve morfotaktik kurallar) ve sözlüktür. Bu tezde morfolojik çözümleyici olarak açık kaynaklı olan Nüve kullanılmıştır. Dilden bağımsız olan Nüve’de, bir dilin tanımlanabilmesi için kökler ve ekler sözlükleri, yazım kuralları ve morfotaktik kurallar dosyası olmak üzere 4 dosyaya ihtiyaç vardır. Nüve’de Kazakça için 145 ekten oluşan ekler sözlüğü, yaklaşık olarak 24000 girdiden oluşan kökler sözlüğü hazırlanmıştır. Üçüncü dosya olarak, XML formatında hazırlanan 26 iki-düzeyleli ortografik (yazım) kural yazılmıştır. Ayrıca simgeler için de (kesme işareti ekleme-silme, alt tire işaretinin silinmesi) kurallar yazılmıştır. Gerekli olan son dosya olan morfotaktik kurallar dosyasında, Sonlu Durum Makinaları kullanılarak Kazakça morfotaktik kodlanmıştır. Daha sonra çizilen isim ve fiil morfotaktik SDM modelleri XML formatında Nüve’de tanımlanmıştır. Nüve’de morfolojik çözümleme yapılırken yazım kurallarına bakılarak geçerli kök ve ek(ler) dizisi bulunur. Bu kök/ek dizisi bulunurken kelimenin kök çözümleri için kökler sözlüğüne ve gövdenin geri kalan kısmı için ekler sözlüğüne bakılır. Ekler sözlüğünde bulunan her ek morfotaktik kurallarına göre o gövdeden/ekten sonra gelebiliyor olması gerekir, yani morfotaktik dosyasında tanımlanmış olmalıdır. Eğer tanımlıysa gövde ve bu ekten oluşan dizi yeni gövde olarak kabul edilir. Bu işlem çözümlenmesi istenilen kelimenin sonuna kadar tekrar edilir. Bu işlem sonucunda, morfotaktik dosyasında tanımlı durumlar arasında bir yol elde edilmişse (final durumuna ulaşılmışsa) bu yol(lar) kelimenin morfolojik çözüm sonucu olarak listelenir.

Kazakçanın morfolojik çözümlemesinin test edilmesi için UD (Universal Dependency) veritabanında Kazakça için hazırlanmış olan veri seti kullanılmıştır. Bu veri setinde her cümle, cümle etiketleri ile birbirinden ayrılmıştır. Daha sonra Nüve’de morfolojik çözümlemesi yapılarak cümleyi oluşturan her kelime/ kelime öbeği için morfolojik çözüm(ler) bulunmuştur. Kazakça ek bakımından zengin bir dil yapısına sahiptir ve bu da morfolojik çözümlerin sayıca fazla çıkmasına sebep olmuştur. Morfolojik çözüm sayısının fazla çıkması çözümleyicinin performansı bakımından avantaj olsa bile hangi çözümün doğru olduğunun bulunması ve seçilmesi problemini (morfolojik belirsizlik) ortaya çıkarmıştır. Nüve’de bulunan morfolojik

çözümlerde öncelikle manuel olarak temizleme işlemi yapılmıştır. İlk 5 beş morfolojik çözüme kadar dikkate alınmıştır. Daha sonra her kelime için bulunan çözümlerden doğru olanı ilk çözüm olarak elle işaretlenmiştir ve varsa diğer çözümler tab ile ayrılarak yazılmıştır. Çözümleme sonrasında elde edilen morfolojik çözümler veri seti toplamda 1073 cümleden, 10000 den fazla kelime girdisinden oluşmaktadır. Bu çözümleme arasında Nüve’de çözümü bulunamayan kelimeler (girdiler) bilinmeyen olarak işaretlenmiştir.

Bir dilin sözdizimsel olarak çözümlenebilmesi için gramer kurallarına (XML formatında) ve morfolojik çözümler dosyasına ihtiyaç vardır. Sözdizimsel çözümleme, morfolojik çözümlemenin sonuçlarını kullanarak cümleyi oluşturan öğeler arasındaki bağlantıya göre cümleyi analiz eder. Kazakça gibi sondan eklemeli diller de cümleyi oluşturan kelimeler birbirine ekler aracılığıyla bağlıdır. Kazakçanın sözdizimsel analizi için Nüve sözdizimsel çözümleyici (Fatih Parser’ın devamı) geliştirilmiştir. Türkçe ve Türkmence sözdizimsel çözümleme Nüve’de hazır olarak bulunmaktadır. Kullanılan sözdizimsel çözümleyicinin kullandığı morfolojik çözümler formatı ile tezin ilk aşamasında kullanılan morfolojik çözümleyicinin çıktı formatı birbirinden farklıdır. Bu farkı ortadan kaldırmak için manuel tag eşleştirilmesi yapılarak sözdizimsel çözümleyicinin kullanabileceği formata dönüştürülmüştür. Sözdizimsel çözümleme için gerekli olan gramer dosyasında yaklaşık olarak 200 CFG kuralı XML formatında yazılmıştır. Bu gramer kuralları ve morfolojik çözümlere bakılarak verilen bir cümlelerin analizi yapılmıştır. Daha sonra cümle analizinin gösterilmesi için çözüm ağacı (parse tree) gösterimi kullanılmıştır. Bir cümle için çizilen her çözüm ağacı doğru olmayabilir. Bu yüzden bu ağaçların doğruluklarının kontrol edilmesi için doğrulama modülüne (parse Tree Validator) ihtiyaç duyulmuştur. Bu modülde tıpkı gramer kuralları gibi kurallar XML formatında yazılmıştır. Çözüm ağaçları arasında doğrulama işleminden sonra geçerli çözüm ağaçları listelenir. Cümleyi oluşturan her kelimenin morfolojik çözüm sayısına bağlı olarak cümlelerin çözüm ağacı sayısı değişir.

Bu tezde, kural tabanlı Kazakça morfolojik ve sözdizimsel analizinin yapılmıştır. Doğal dillerin (özellikle Kazakça gibi sondan eklemeli dillerin) gerek ekler gerekse de kelimeler bakımından çok kapsamlı olmasından dolayı, her durumu ele alabilecek kuralların yazılması zordur. Her istisnai durum için kurallar yazılmış olsa da bazı durumların çözümü bulunamamıştır.

Sözdizimsel analiz aşamasında, manuel olarak hazırlanmış Kazakça veri seti üzerinde testler yapılmıştır. Morfolojik çözümleme ve sözdizimsel çözümleme birbirinden ayrı olarak

çalışmaktadır. Bunun geliştirilerek tek bir çözümleyici (hem morfolojik çözümleyici hem de sözdizimsel çözümleyici) olarak birleştirilmesi daha iyi olacaktır. Yani Nüve''ye verilen bir girdinin (metnin) otomatik olarak morfolojik çözümlerinin bulunduktan sonra sözdizimsel analizi yapılması daha kullanışlı olacaktır.

Bu tezin amaçlarından biri de yapılan bu çalışmaların web ortamına taşıyarak herkese açık hale getirmektir. Bu doğrultuda <https://nuvestudio.com/> web sayfası çalışma grubumuz tarafından geliştirilmiştir. Bu web ara yüzünde Kazakça, Türkçe, Özbekçe dilleri için morfolojik üretim (Ortography sekmesinde) ve çözümleme (Morphotactics sekmesinde) hazır hale getirilmiştir. Bu web sayfasına sözdizimsel çözümleme ve diller arasındaki çevirinin de eklenmesiyle çevrimiçi morfolojik, sentaktik çözümleyici ve makine çevirisi sistemi kullanıma açılabilir.

KAYNAKLAR

- ABDUSSAITOVA, A. & AMANGELDIYEVA, A. Normalization of Kazakh Texts. Proceedings of the Student Research Workshop Associated with RANLP 2019, 2019. 1-6.
- AKIN, A. A. & AKIN, M. D. 2007. Zemberek, an open source nlp framework for turkic languages. *Structure*, 10, 1-5.
- AKYÜREK, E., DAYANIK, E. & YURET, D. 2019. Morphological analysis using a sequence decoder. *Transactions of the Association for Computational Linguistics*, 7, 567-579.
- ALTENBEK, G. & WANG, X.-L. Kazakh segmentation system of inflectional affixes. CIPS-SIGHAN Joint Conference on Chinese Language Processing, 2010.
- ALTINTAŞ, K. 2001. *Turkish to Crimean Tatar machine translation system*. Bilkent University.
- ASSYLBEKOV, Z., WASHINGTON, J., TYERS, F., NURKAS, A., SUNDETOVA, A., KARIBAYEVA, A., ABDUALI, B. & AMIROVA, D. 2016. A free/open-source hybrid morphological disambiguation tool for Kazakh.
- BABÜROĞLU, B., TEKEREK, A. & TEKEREK, M. 2019. Türkçe için derin öğrenme tabanlı doğal dil işleme modeli geliştirilmesi. Web: <https://arxiv.org/ftp/arxiv/papers/1905/1905.05699.pdf>, 16, 2019.
- BEKMANOVA, G., SHARIPBAY, A., ALTENBEK, G., ADALI, E., ZHETKENBAY, L., KAMANUR, U. & ZULKHAZHAY, A. A Uniform Morphological Analyzer for the Kazakh and Turkish Languages. AIST (Supplement), 2017. 20-30.
- BIRAY, N. 2003. Türkiye Türkçesi ile Kazak Türkçesindeki cümlelerin çeşitleri bakımından karşılaştırılması üzerine bir deneme. *Selçuk Üniversitesi Türkiyat Araştırmaları Dergisi*, 281-307.
- BIRAY, N., AYAN, E. & ERCILASUN, G. K. 2015. *Çağdaş Kazak Türkçesi Ses - Şekil - Cümle Bilgisi - Metinler*, BİLGE KÜLTÜR SANAT.
- BRILL, E. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational linguistics*, 21, 543-565.
- ÇÖLTEKİN, Ç. A Freely Available Morphological Analyzer for Turkish. LREC, 2010. 19-28.
- ÇÖLTEKİN, Ç. 2013. TRmorph: A morphological analyzer for Turkish.
- DAYANIK, E., AKYÜREK, E. & YURET, D. 2018. Morphnet: A sequence-to-sequence model that combines morphological analysis and disambiguation. *arXiv preprint arXiv:1805.07946*.
- DAYBELGE, T. & CICEKLI, I. A rule-based morphological disambiguator for Turkish. Proceedings of Recent Advances in Natural Language Processing, 2007. 145-149.
- DELIBAS, A. 2008. *Doğal dil işleme ile Türkçe yazım hatalarının denetlenmesi*. Fen Bilimleri Enstitüsü.
- ERYİĞİT, G. 2012. Biçimbilimsel çözümleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.
- ERYİĞİT, G. & ADALI, E. An affix stripping morphological analyzer for Turkish. Proceedings of the IASTED international conference artificial intelligence and applications, 2004. 299-304.
- FATULLAYEV, R., ABBASOV, A. & FATULLAYEV, A. 2008. Dilmanc is the 1st MT system for Azerbaijani. *Proc. of SLTC-08, Stockholm, Sweden*, 63-64.
- GILMULLIN, R. 2008. The Tatar-Turkish machine translation based on the two-level morphological analyzer. *Interactive Systems and Technologies: The Problems of Human-Computer Interaction*, 179-186.

- GLOROT, X. & BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010. JMLR Workshop and Conference Proceedings, 249-256.
- GOLDBERG, Y. 2016. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57, 345-420.
- GÖKIRMAK, M., TYERS, F. & WASHINGTON, J. Machine Translation for Crimean Tatar to Turkish. Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages, 2019. 24-31.
- HAKKANI-TÜR, D. Z., OFLAZER, K. & TÜR, G. 2002. Statistical morphological disambiguation for agglutinative languages. *Computers and the Humanities*, 36, 381-410.
- HAMZAOGLU, İ. 1993. *Machine translation from Turkish to other Turkic languages and an implementation for the Azeri languages*. Master of Science, Bogazici University.
- HULDEN, M. Foma: a finite-state compiler and library. Proceedings of the Demonstrations Session at EACL 2009, 2009. 29-32.
- KAIRAKBAY, B. M. & ZAURBEKOV, D. L. Finite state approach to the Kazakh nominal paradigm. Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing, 2013. 108-112.
- KARAHAN, L. 2008. *Türkçe'de Söz Dizimi*, Ankara, Akçağ.
- KAYA, Y. & ERTUĞRUL, Ö. F. 2016. Doküman dili tanıma için yeni bir öznitelik çıkarım yaklaşımı: İkili desenler. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 31.
- KAYA, Y., ERTUĞRUL, Ö. F. & TEKİN, R. 2015. Doküman dili tanıma için ikili örüntüler tabanlı yeni bir yaklaşım. *Akademik Bilişim, Eskişehir*.
- KAYA, Y. & ÖZDEMİR, C. 2019. Yeni öznitelik çıkarım yöntemleri ile duygu analizi.
- KESSIKBAYEVA, G. & CICEKLI, I. Rule based morphological analyzer of Kazakh language. Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM, 2014. 46-54.
- KOSKENNIEMI, K. 1983. *Two-level morphology: A general computational model for word-form recognition and production*, University of Helsinki, Department of General Linguistics Helsinki, Finland.
- KOZHIRBAYEV, Z. & YESSENBAYEV, Z. Kazakh Text Normalization using Machine Translation Approaches. CEUR Workshop Proceedings, 2020. CEUR-WS, 115-122.
- MAIMAITI, M., WUMAIER, A., ABIDEREXITI, K. & YIBULAYIN, T. 2017. Bidirectional long short-term memory network with a conditional random field layer for Uyghur part-of-speech tagging. *Information*, 8, 157.
- MAKAZHANOV, A., MAKHAMBETOV, O., SABYRGALIYEV, I. & YESSENBAYEV, Z. Spelling correction for kazakh. International Conference on Intelligent Text Processing and Computational Linguistics, 2014. Springer, 533-541.
- MAKAZHANOV, A., SULTANGAZINA, A., MAKHAMBETOV, O. & YESSENBAYEV, Z. Syntactic annotation of kazakh: Following the universal dependencies guidelines. a report. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE" TURKIC LANGUAGES PROCESSING" TurkLang-2015, 2015. 338-350.
- MAKHAMBETOV, O., MAKAZHANOV, A., SABYRGALIYEV, I. & YESSENBAYEV, Z. Data-driven morphological analysis and disambiguation for kazakh. International Conference on Intelligent Text Processing and Computational Linguistics, 2015. Springer, 151-163.
- MAKHAMBETOV, O., MAKAZHANOV, A., YESSENBAYEV, Z., MATKARIMOV, B., SABYRGALIYEV, I. & SHARAFUDINOV, A. Assembling the kazakh language

- corpus. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013. 1022-1031.
- MAKHAMBETOV, O., MAKAZHANOV, A., YESSENBAYEV, Z., SABYRGALIYEV, I. & SHARAFUDINOV, A. 2014. Towards a data-driven morphological analysis of Kazakh language. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 7, 31-36.
- OFLAZER, K. 1994. Two-level description of Turkish morphology. *Literary and linguistic computing*, 9, 137-148.
- OFLAZER, K. 2016. Türkçe ve Doğal Dil İşleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.
- OFLAZER, K., GÖÇMEN, E. & BOZŞAHİN, C. 1994. An outline of Turkish morphology. *Report to NATO Science Division Sfs III (TU-LANGUAGE)*, Brussels.
- OFLAZER, K. & TUR, G. 1996. Combining hand-crafted rules and unsupervised learning in constraint-based morphological disambiguation. *arXiv preprint cmp-lg/9604001*.
- OFLAZER, K. & TUR, G. 1997. Morphological disambiguation by voting constraints. *arXiv preprint cmp-lg/9704011*.
- ORHUN, M., TANTUĞ, A. C. & ADALI, E. Morphological disambiguation rules for Uyghur language. 2010 IEEE International Conference on Software Engineering and Service Sciences, 2010. IEEE, 542-546.
- SALIMZYANOV, I., WASHINGTON, J. & TYERS, F. A free/open-source Kazakh-Tatar machine translation system. Proceedings of Machine Translation Summit XIV: Papers, 2013.
- SANLI, T. 2018. *Kırım Tatarcasının biçimbilimsel çözümlemesi ve Kırım Tatarcası-Türkçe biçimbilimsel makina çevirisi sistemi*. Yüksek Lisans, İstanbul Üniversitesi Fen Bilimleri Enstitüsü.
- SEVİM, S. 2003. *Kazak Türkçesi ile Türkiye Türkçesinin sentaksının karşılaştırılması*. Master, Pamukkale Üniversitesi.
- SUNDETOVA, A., KARIBAYEVA, A. & TUKEYEV, U. 2014. Structural Transfer Rules for Kazakh-to-English Machine Translation in The Free/Open-Source Platform Apertium. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 7, 48-53.
- ŞEKER, Ş. E. 2015. Doğal Dil İşleme (Natural Language Processing). *YBS Ansiklopedi*, 2, 14-22.
- TANTUĞ, A. C. & ADALI, E. 2018. Machine translation between Turkic languages. *Turkish Natural Language Processing*. Springer.
- TOLEU, A., TOLEGEN, G. & MAKAZHANOV, A. Character-aware neural morphological disambiguation. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2017. 666-671.
- TUKEYEV, U., MELBY, A. & ZHUMANOV ZH, M. Models and algorithms of translation of the Kazakh language sentences into English language with use of link grammar and the statistical approach. Proc. of IV Congress of the Turkic World Math. Society, 2011. 1-3.
- TYERS, F. M. & WASHINGTON, J. Towards a Free/Open-source Universaldependency Treebank for Kazakh. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE" TURKIC LANGUAGES PROCESSING" TurkLang-2015, 2015. 276-289.
- YINER, Z. 2016. *Kırgızca'nın biçimbilimsel çözümlemesi ve Kırgızca-Türkçe biçimbilimsel makina çevirisi sistemi*. Yüksek Lisans, İstanbul Üniversitesi Fen Bilimleri Enstitüsü.

- YURET, D. & TÜRE, F. Learning morphological disambiguation rules for Turkish. Proceedings of the Human Language Technology Conference of the NAACL, Main Conference, 2006. 328-334.
- ZAFER, H. R. 2011. *A generic syntactic parser for Turkic languages*. Master of Science, Fatih University The Graduate Institute of Sciences and Engineering.

