Hacettepe University Graduate School of Social Sciences

Department of English Linguistics

A COMPUTER MEDIATED ANALYSIS OF NEOLOGISMS USED BY TURKISH SPEAKERS ON X SOCIAL MEDIA PLATFORM

Berkay EROL

Master's Thesis

Ankara, 2024

A COMPUTER MEDIATED ANALYSIS OF NEOLOGISMS USED BY TURKISH SPEAKERS ON X SOCIAL MEDIA PLATFORM

Berkay EROL

Hacettepe University Graduate School of Social Sciences

Department of English Linguistics

Master's Thesis

Ankara, 2024

ACCEPTANCE AND APPROVAL

The jury finds that Berkay EROL has on the date of September 3rd 2024 succesfully passed defense examination and approves his Master's Thesis tittled "A Computer Mediated Analysis of Neologisms Used by Turkish Speakers on X Social Media Platform."

Assoc. Prof. Dr. Gökçen HASTÜRKOĞLU (Jury President)

Assoc. Prof. Dr. Zeynep DOYURAN (Advisor)

Assoc. Prof. Dr. Suhan AKINCI OKTAY

Assoc. Prof. Dr. Yeliz DEMİR VAN SCHEPPINGEN

Assit. Prof. Dr. Taylan AKAL

I agree that the signatures above belong to the faculty members listed.

Prof. Dr. Uğur ÖMÜRGÖNÜLŞEN

Graduate School Director

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan ***"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ….. ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

……/……/2024

Berkay EROL

[1] *"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*

(1) Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulu*** *iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulunun*** *gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, ***tezin yapıldığı kurum*** *tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise,* ***ilgili kurum ve kuruluşun önerisi*** *ile* ***enstitü*** *veya* ***fakültenin*** *uygun görüşü üzerine* ***üniversite yönetim kurulu*** *tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

****** Tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulu tarafından karar verilir.***

ETİK BEYAN

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, **Doç. Dr. Zeynep DOYURAN** danışmanlığında tarafımdan üretildiğini ve Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Yazım Yönergesine göre yazıldığını beyan ederim.

Berkay EROL

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Assoc. Prof. Dr. Zeynep DOYURAN, for her reliable guidance and constant encouragement. Her constant support and feedback guided me through this thesis.

I also thank the members of my thesis jury, Assoc. Prof. Dr. Gökçen HASTÜRKOĞLU, Assoc. Prof. Dr. Suha AKINCI OKTAY, Assoc. Prof. Dr. Yeliz DEMİR VAN SCHEPPINGEN, and Assist. Prof. Dr. Taylan AKAL, for their constructive criticism and guidance during the defence examination.

Without a doubt, none of this would be possible without the support of my family, specifically my mother, Ayla EROL, who offered me her unconditional love and encouragement throughout the whole journey. I would also like to thank my brother Burak EROL for being there for me whenever I needed an ear to listen. I am indebted to my maternal aunt Özden GÖNÜLTAŞ and my father, Bülent EROL, for supporting my dreams and goals since my childhood. I love you all deeply. I could not achieve my goals without you.

Last but not least, I would like to thank all my dear friends who helped me through this journey, especially Işıl İrem KELEŞ, since she was always there for me through the thick and thin, and I thank her for all the laughs we shared and study nights we held together. I would like to express my sincere gratitude to Buğrahan HALICI for being a very patient and supportive tutor. Through his efforts, I could learn coding with Python and made this study possible. I am grateful for all the support my friends showed to me during my journey.

ABSTRACT

EROL, Berkay. *A Computer Mediated Analysis of Neologisms Used by Turkish Speakers on X Social Media Platform*, Master's Thesis, Ankara, 2024.

Neologism refers to newly emerged words and expressions or new definitions for existing words or expressions. Neologisms are an essential part of lexicography and etymology studies since new words constantly emerge in lexicons due to the dynamic nature of languages. There are recent efforts to compile neologisms in Turkish for lexicology studies. In addition, social media application programming interfaces (APIs) have become a recent trend in compiling massive quantities of data. This study facilitates X API to gather Tweets from X to compile a corpus and analyse and categorise neologisms used on social media. The corpus is compiled from Tweets sent from Turkey in Turkish between 01.01.2023 and 31.12.203. The corpus amounts to 327.262 Tweets with a total word count of 2.463.075. These Tweets are then tokenised by using *TRNLP* for further analysis. The tokenised entries are morphologically analysed via *TrMorph* to account for lemmatisation. The resulting data was analysed to identify Turkish neologisms most prevalent on the social media platform X. The selected neologisms account for lemmatisation and semantic shifting. Definitions of all neologisms found in the study are explained with examples. These neologisms are then analysed based on five categories: their frequency in the data set, function, coinage, formation process, and source. The study finds that neologisms used by Turkish speakers on social media are primarily expressive in function. Furthermore, the study finds that all the neologisms formed through borrowings were directly taken from English. The study also presents five new neologism formation methods for the Turkish language: blending, hypocoristic neologisms, hybrid neologisms, phono-semantic shifts, and phraseology. Hypocoristic neologisms are proposed as a new neology formation method unique to Turkish.

Keywords

Neologism, Language Change, Social Media, X, Tweets, Turkish Language

ÖZET

EROL, Berkay. *Sosyal Medya Platformu X'te Türkçe Konuşucuların Kullandıkları Neolojizmlerin Bilgisayar Destekli İncelenmesi*, Yüksek Lisans Tezi, Ankara, 2024.

Neolojizm, yeni ortaya çıkan kelime ve ifadeleri veya mevcut kelime ya da ifadelere getirilen yeni tanımları ifade eder. Neolojizmler, dillerin dinamik yapısı nedeniyle sözlüklerde sürekli olarak yeni kelimeler ortaya çıktığı için sözlükbilim ve etimoloji çalışmalarının önemli bir parçasıdır. Türkçede neolojizmleri derlemek için yapılan son çalışmalar, sözlükbilim alanında önemli bir yer tutmaktadır. Buna ek olarak, sosyal medya uygulama programlama arayüzleri (API'ler), büyük miktarda veri derlemek için son zamanlarda yaygın bir trend haline gelmiştir. Bu çalışma, X API'sini kullanarak X'ten Tweet toplayıp bir derlem oluşturmayı ve sosyal medyada kullanılan neolojizmleri analiz edip sınıflandırmayı amaçlamaktadır. Derlem, 01.01.2023 ile 31.12.2023 tarihleri arasında Türkiye'den atılan Türkçe Tweetlerden derlenmiştir. Derlem, toplamda 2.463.075 kelimeden oluşan 327.262 Tweet içermektedir. Bu Tweetler, daha detaylı analiz edilmek üzere *TRNLP* kullanılarak tokenleştirilmiştir. Tokenleştirilen veriler, lematizasyon amacıyla *TrMorph* kullanılarak morfolojik bakımdan analiz edilmiştir. Elde edilen veriler, sosyal medya platformu X üzerinde en yaygın kullanılan Türkçe neolojizmleri belirlemek amacıyla incelenmiştir. Seçilen neolojjzmler lematizasyon ve anlam kayması göz önünde bulundurularak ele alınmıştır. Çalışmada bulunan tüm neolojizmlerin örneklerle beraber tanımı yapılmıştır. Bu neolojizmler daha sonra dört kategoriye göre analiz edilmiştir: işlevleri, türetilme biçimleri, oluşum süreçleri ve kaynakları. Her bir kategori için sıklık analizi yapılmıştır. Çalışma, Türkçe konuşan sosyal medya kullanıcıları tarafından kullanılan neolojizmlerin çoğunlukla ifadeye dayalı işlevlerde olduğunu ortaya koymaktadır. Ayrıca, çalışmada tespit edilen tüm ödünç alınma yöntemiyle oluşturulan neolojizmlerin doğrudan İngilizceden alındığı sonucuna ulaşılmıştır. Çalışma, Türk dili için beş yeni neolojizm oluşum yöntemi sunmaktadır: harmanlama, küçültme neolojizmleri, hibrit neolojizmler, ses-anlam kaymaları ve öbeksel neolojizmler. Küçültme neolojizmleri, tamamen yeni ve Türkçeye özgü bir neolojizm oluşum yöntemi olarak önerilmektedir.

Anahtar Sözcükler

Neolojizm, Dil Değişimi, Sosyal Medya, X, Tweet, Türk Dili

TABLE OF CONTENTS

INDEX OF ABBREVIATIONS

API **:** Application Programming Interface

TDK **:** Türk Dil Kurumu

TRNLP **:** Turkish Natural Language Processing

INDEX OF TABLES

INTRODUCTION

This thesis is concerned with neologisms in the Turkish language on the social media platform X, previously known as Twitter, across Türkiye. It aims to identify the usage of neologisms proposed by previous studies and their usage in X. It also identifies new neologisms based on Tweets published on the same platform. Furthermore, it defines the selected neologisms used on social media. The word “neologism” comes from the blending of two Greek words; “neo” (new) and “logos” (word). Neologism refers to newly emerged words and expressions or new definitions for existing words or expressions. Identifying a neologisms is considered a problematic issue (Boulanger, 2010). They may be completely new words or new definitions with existing words. Neologism may spring from intentional acts of linguistic creativity, where individuals or groups create new ways to encapsulate novel ideas or phenomena. However, it may also occur organically via linguistic processes such as blending, derivation, or coinage. These new words may spread around a more closed sub-culture, such as a game-related term spreading among young gamers, or it may become widespread if the neology is related to a phenomenon that concerns society more broadly. Historically, neologisms refer to newly emerged concepts such as the word “television” or “internet” after their invention. However, slang expressions can also become long-lasting neologisms and enter standardised dictionaries (Brittanica, 2024). There are scientific methods to identify neologisms. The first is using corpus-based approaches to analyse the frequency and meaning of neologisms. This is very much like a dictionary inclusion criteria that lexicographers use. Merriam-Webster Dictionary states that a new word is incorporated into the dictionary based on frequency, widespread use, and meaningful use. Frequency criteria are based on how frequently a word is used. Widespread use refers to how many speakers use the word, a criterion used to eliminate words only used in professional contexts, as specialised dictionaries are a better medium to define such words. Finally, meaningful use criteria require a word to be used in a way that it describes. One way to reflect and exemplify

this phenomenon would be through the word “muvaffakiyetsizleştiricileştiriveremeyebileceklerimizdenmişsinizcesine” the famous word in Turkish due to its structure and length. This is the longest possible meaningful word in Turkish. However, it is nearly never used in its intended meaning; instead, it is used to exemplify the long words a speaker can produce in Turkish. Thus, such words are excluded from dictionaries even if they satisfy other criteria. The distinction between corpus-based neologism identification and lexicographical methods for new dictionary entries is that dictionaries such as Merriam-Webster study the frequency of articles, books, and speech. However, they overlook more informal communication media, such as social media. However, corpus-based neologism identification also tries to identify words that are not in dictionaries or not used in the way they are described in dictionaries.

Due to its nature, the digital age helps spread information on a much faster chassis. Accessing information is a fast and easy process, and digital platforms such as social media significantly quicken the process of spreading. Constant exposure to slang, different linguistic variations, neologisms, regional dialects, etc., also facilitates the integration of non-standard linguistic features. According to Dannet and Herring (2007), online platforms allow users to interact with each other asynchronously without the limitations of geographical and temporal boundaries, which facilitates the formation of global communities and the exchange of ideas. Therefore, the spread of a neologism and its probability of becoming an official dictionary entry is not constrained as it was in the past.

According to Tahiroğlu (2014), neologism mainly occurs due to socio-economic, socio-cultural, and psychological reasons rather than linguistic reasons. Even when a neologism spreads fast, it does not necessarily mean it would be used by every speaker on the same frequency because the socio-economic, socio-cultural, or psychological reasons they formed out of will still not be relevant for every speaker. The use of informal lexicol and variant spellings on X continues to reflect patterns of variation in spoken language that align with geographical

and demographic distribution (Eisenstein et al., 2014; Huang et al., 2016). This shows that social identities are still dominant in how users use the language.

Linguistic variables are the different ways any linguistic element can be realised, and any different realisation is considered a new variant. On the other hand, sociolinguistic variables can be associated with different social identities as the base reason. Shoemark (2020), analysing the relative frequencies of variables, reveals how people choose to refer to things rather than what they are referring to.

This thesis presents data collected from the social media platform X using Python through the platform's official API. The collected data are based on the neologisms in the Turkish language catalogued by Tahiroğlu (2014), Çokol (2020), and Bilginsoy (2019). Furthermore, the study uses a frequency analysis of Tweets collected from the platform to identify new neologisms on the platform and give definitions to them. In this thesis, neologisms refer to lexical items not available in the official dictionary of the Turkish Language Association. User bios, short and limited sections users can fill in to describe themselves, and their tweets are included in the statistical analysis. Tweets that do not contain the user’s location or are Tweeted outside Türkiye are excluded from the study analysis.

BACKGROUND

In this section, notions of understanding neologisms and categorising them will be explained in greater detail. First, neologisms and what constitutes a neologism will be explained in greater detail. Second, the types of word formation processes for neologisms will be explained. Finally, how social media is used to gather data for the study will be laid out.

NEOLOGISMS

According to the Cambridge Online Dictionary, neologisms are defined as “a new word or expression, or a new meaning for an existing word”. Thus, neologisms not only include existing lexical forms with new meanings (which is also called semantic shifting) (Newmark, 1988; Jamet, 2018; Jamet & Terry, 2018) but also new forms and novel meanings (Ulanova, 2014; Cook, 2010; Cook, 2018; Rets, 2014). The systematic and formal study of neologisms started with the first neologism dictionary published by Dwight Bolinger in the early 1930s.

Neologisms may be born out of artistic creativity, as seen in literature. One of the most famous examples of such creativity would be the word “Jabberwocky” from the book titled with the same name, written by Lewis Caroll. Jabberwocky is a fictional creature that the protagonist fights in the book. The word is born out of the artistic creativity of the author. Another way a neologism may be born is when an existing notion lacks a term to identify it or when speakers are unaware of the existing term for it. Primarily, scientific or professional jargon is rich with such neologisms. An example of a neologism created this way would be the word for “computer”. When it was a new invention, it required a term to capture the notion of an electronic device that could compute information and mathematics. Thus, the inventors came up with the name computer. However, neologisms could also born organically. Speakers may identify a notion or concept with no direct term to refer to and develop their term organically. One example of such a word would be Pollyanna, a character in Eleanor H. Porter’s book titled ‘Pollyanna’. Pollyanna is a character in the book but is also used as an adjective to define overly optimistic people.

Identifying what is new and what is not can be problematic, as stated by researchers such as Boulanger (2010). However, there is an approach to determine if a word fits the criteria of neologisms: inclusion in dictionaries. Cabré (1993) states that neologisms are words not yet included in dictionaries. This approach is used by many others, such as Humble (2006), Jamer and Terry

(2016) and Tahiroğlu et al. (2014), to provide a baseline for identifying neologisms. Guerra (2016) states that this method is currently the most accepted approach to neologisms. This approach partially eliminates the need to pinpoint the exact date of the creation of a given neologism. Indeed, in some circumstances, identifying the exact date of when a neologism was created can be done. If neologism is born out of artistic creativity, such as the example of 'Jabberwocky', its exact first usage could be pinpointed in time by simply looking at the publishment date of the book. If it is a term for a new notion, such as the invention of computers or aeroplanes and the need to identify these concepts, the exact date or at least year can be pinpointed by simply looking at the date of invention for the notion. Suppose the neologism emerged to refer to a concept newly introduced to the language, such as the COVID-19 pandemic and related neologisms that emerged out of it. In that case, the exact year can be pinpointed easily. These types of neologisms can be tracked due to the formal or global nature of their coming into existence. However, this is not true for every neologism. An example is the name Pollyanna, which started to be used as an adjective to define overtly optimistic people. It is hard to track back on time unless it happened to be catalogued by linguists. However, cataloguing a neologism that emerged in small groups or a neologism that emerged to be used in informal settings may prove a bigger challenge than lexicologists could tackle. It is hard to scourge through the World Wide Web to pinpoint the first use of a neologism. Even if it could be pinpointed, there would be no evidence to suggest that it was not in verbal speech before being used on the internet. It would be nearly impossible to track every utterance that comes out of a language's speakers, catalogue them, and check if a neologism is uttered. Furthermore, if the neologism emerged out of a private segment of a social media page (i.e. private Instagram accounts, private Facebook groups, private X accounts, private Tumblr accounts, private chat application conversations, etc.), both ethically and technically, it would not be feasible to access that data. Ethically, a researcher would need to have consent from every relevant party for a group or both parties for private conversations to gather the data. It would be nearly impossible to get consent from every user of every social media platform to scrutinise private

segments. However, even with consent from relevant parties, this would still technically prove impossible as the methods to gather mass data from social media websites through APIs do not allow access to such private groups and profiles and private conversations, be it a chat group or a bilateral chat are encrypted and not visible to API searches or any third party application. Considering neologisms can originate from small circles and get spread through a long period, exactly pinpointing the first emergence of a naturally occurring neologism would prove impossible. Therefore, considering the dictionary entries as a base allows a concrete method to identify neologisms as a widespread enough word would get into dictionaries as per the method of dictionary additions stated by major dictionaries such as Merriam Webster lays out criteria for how new entries get added to their dictionary. However, the most prominent Turkish dictionary, the Turkish Dictionary published by Türk Dil Kurumu (TDK), does not clearly state how new words are added to the dictionary. Three methods can be somewhat identified through research on how TDK adds new entries to its dictionary. The first one is Tahiroğlu et al.'s (2014) project that identifies neologisms and their frequency through online newspapers and creates a database for TDK. The second is a small statement TDK made on their website for frequently asked questions. TDK specifically answer to the question "Yeni bir kelime buldum/uydurdum. Sözlüğe alınması mümkün mü?" (I have created/made up a new word. Is it possible to add it to the dictionary?) as "Türk Dil Kurumu, dilimizin söz varlığının belirleyicisi değil bilimsel anlamda derleyicisi ve sınıflayıcısıdır. Kişilerin ürettiği veya türettiği sözler, sözlüğe alınamaz. Bir kelimenin sözlüğe alınması, o kelimenin halk tarafından benimsenip kullanılmasına ve dilimize yerleşmesine bağlıdır" (The Turkish Language Association is not the determiner of our language's vocabulary but its compiler and classifier in a scientific sense. Words created or derived by individuals cannot be included in the dictionary. Including a word in the dictionary depends on whether it is adopted and used by the public and has become established in our language.". Therefore, we can conclude that TDK did include new entries to dictionaries when they started to get used frequently by the general public. The third method we can see that TDK employs for new dictionary entries is through

a suggestion form they put up on sozluk.gov.tr. This form can be filled out to find new lexical units for borrowed words. TDK states that they add new entries if the public frequently uses them. However, any additional criteria that they may have are not explicitly stated. Still, this provides a baseline for the thesis as neologisms frequently used by the public will be added to TDK's dictionary, according to TDK's statement.

SOCIAL MEDIA PLATFORMS

Social media websites are places people use every day to communicate, whether they chat, get into discussions, share news, or share their stories. Social media's whole concept is based on human interaction, which happens mostly through written communication. Social media sites record a massive amount of communication data each day, and openly public interactions on social media platforms can be collected through their APIs.

Social media is being used by researchers such as Grieve et al. (2017), Pinto et al. (2020), Monderin (2021) to find emerging words of a language. Social media APIs provide large quantities of data to researchers; since frequency is an essential factor in such research, they facilitate a way of doing research that cannot be done with traditional methods with relatively low costs and fieldwork. The aforementioned nature of social media sites creates an ethical way of obtaining mass and naturally occurring data.

Social media platforms with public APIs offer a significant volume of spontaneous and informal language data from a diverse user base. However, the most significant advantage of collecting data from a social media platform with API is the ability to completely eliminate the observer's paradox. According to Labov (1978), "the aim of linguistic research in the community must be to find out how

people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation." (p. 209). When data are gathered through API, the conditions created by the observer's paradox are eliminated. Social media APIs allow the researcher to observe naturally occurring conversations between users through passive, non-intrusive means. The researcher does not need to facilitate conversation, interview, or give handouts. Another benefit is being able to reach significant volumes of data with ease, eliminating the need to rely on small sample sizes and increasing the generalizability of the research. Another critical factor is being able to gather naturalistic data since social media platforms allow users to interact socially with each other without relying on artificially generated settings or controlled interviews. It is a naturalistic, genuine, and real-time interaction between the users. Moreover, data gathered through social media APIs provide greater anonymity to users since it allows the researcher to anonymise and aggregate the whole data, creating safeguards for individuals and communities.

One of the main benefits of using X API is the metadata accompanying the tweets, such as time stamps, emojis, locations, photographs, and user names. These data sets allow the researcher to create a fine-grained analysis of socio-cultural variables with linguistic variations or changes.

COMPARISON WITH OTHER WRITTEN MEDIA

The main difference between written and spoken mediums can be argued as the persistency of their natures. However, this argument is primarily flawed in the digital world. Speech can be recorded and stored similarly to written mediums, and written mediums can disappear just as quickly as in old forums or blogs. Boyd (2008) suggests four elements in identifying 'networked publics': persistence, replicability, searchability, and scalability.

In terms of persistence, the actual difference between written mediums and speech is the factor of naturality and its scalability regarding naturality. People

generally do not tend to record their daily spoken interactions, but by being digital, social media records every conversation within the platform. If speech is recorded to be persistent, it usually means the process was conducted in a controlled environment by a researcher, which takes away from the naturality of the process and perhaps even creates Oberver's paradox. One way to circumvent this would be to gather voice recordings in public areas without choosing participants. However, this puts forward two different problems. The first one is that this is not scalable as speech needs to be gathered over a longer time than gathering written data from an API, and it would even take longer if the research requires data from a broad geographical region or perhaps even different regions.

Data from radio or TV shows can overcome the scalability issue. However, this method eventually causes issues regarding the naturality of the data. Media organisations and government agencies control and monitor radio and TV channels, creating an unideal environment that may limit the data's naturality. On the other hand, social media is entirely decentralised and provides more natural data for the researcher.

Searchability may be limited in speech data based on transcripts, annotations, or metadata. It especially becomes problematic when the data has no transcription. However, social media data can always be easily searchable with the built-in tools of the APIs or sometimes even the application itself.

Replicability goes hand in hand with searchability. Replicability, in this sense, refers to the scale at which other researchers can replicate a search query and its results. Since every public interaction on social media platforms is stored and saved indefinitely, any search queries conducted through them can be easily replicated by other researchers and yield the same result if search parameters are completely aligned. It should be noted that this may not hold for randomly sampled data. However, randomly sampled data still could be retrieved as it is from social media platforms.

Social media platforms offer the most optimal way of gathering vast data based on the abovementioned elements. Social media offers the most extensive scale among every written media platform. The problem with other written media is that they usually publish or share a very limited amount of text, which is often written by the same writers. Comparing this with social media platforms, some of which have over a billion registered users, they do not get the necessary data influx both in terms of the number of speakers and the amount of written speech data they offer. Social media platforms allow billions of interactions due to their design nature and number of users. Regarding the number of speakers and the amount of written speech, It would be impossible to gather the same data with a similar speaker variation from other written media.

All written media are persistent in nature as the interactions are recorded and stored as their nature. However, not every written media would offer interactions between individuals or allow an informal discourse. For example, newspapers are very persistent, and digital newspapers can provide a means to compile vast data. However, the language newspapers use would be highly regulated and formal, with minimal interactions between individuals. Books, magazines, journals, etc., share the same problems, which would prove an unideal environment to identify neologisms. Social media platforms are the only written media that allows spontaneous and natural interactions between individuals.

The searchability of social media platforms is rivalled by other written media as well. Digital newspapers allow us to search keywords, specific articles, and specific authors and gather data through APIs. Thus, social media and digital newspapers not only allow one to search anything specific manually but also utilise APIs to allow access to compile vast quantities of data for researchers. There are no other written media that offer the same flexibility and accessibility to researchers.

Digital newspapers match the replicability of search queries conducted on social media platforms. Since both media allow API queries and manual search, any

query conducted by a researcher could be easily replicated by others. Other written media platforms do offer the same amount of replicability, but they cannot offer the same accessibility due to the lack of APIs.

Social media platforms are not unique in what they offer, as other written media can offer the same replicability, same searchability, or same persistency. However, no other written media offers data on a similar scale, and no other written media offers everything that social media does. What makes social media platforms uniquely suitable for studying language is how these four elements combine to create an ideal environment in which a very diverse and vast user base spontaneously and naturally interacts with each other.

STATEMENT OF THE PROBLEM

Neologisms are an integral part of lexicology and etymology studies. Identifying neologisms and cataloguing them provides a base for lexicography efforts to identify possible new words to be given a dictionary entry. It furthermore helps with searchability, which may be limited in speech data based on transcripts, annotations, or metadata. It especially becomes problematic when the data has no transcription. However, social media data can always be easily searchable with the built-in tools of the APIs or sometimes even the application itself.

SIGNIFICANCE OF THE STUDY

The significance of this study lies in its exploration of the neologisms within the Turkish language as manifested on the social media platform X (formerly known as Twitter). This research contributes to understanding how new lexical items emerge and how they are formed in Turkish. It also provides a valuable background for future language variation and change research. The digital age has revolutionised the way information is disseminated, and social media platforms, in particular, play a pivotal role in the rapid transmission of linguistic

innovations. Thus, examining neologisms on a platform as widely used as X provides valuable insights into contemporary linguistic trends and socio-cultural dynamics.

One of this study's primary contributions is its ability to shed light on new lexical items that may enter dictionaries. With its diverse user base and spontaneous interactions, social media presents a unique case for studying how new words and expressions proliferate in informal use. By analysing the possible neologisms, this research highlights the informal part of the Turkish language as it is used among social media users.

Furthermore, this study addresses a significant gap in the existing literature on Turkish linguistics. While there has been considerable research on traditional dialectology and language change, there is a paucity of studies focusing on digital communication and its impact on language. This research, therefore, not only contributes to the field of linguistics but also intersects with digital humanities, offering a contemporary perspective on language evolution in the digital era. The study provides empirical evidence that complements and expands upon previous research by utilising data from X. It offers a dynamic view of language as it is actively used and modified in real-time.

The methodological approach of this study is also of considerable significance. Employing the official API of X for data collection ensures that the research is grounded in authentic and current linguistic data. Integrating user-generated content, including tweets and user bios, enables a comprehensive analysis of language use. This approach allows for identifying neologisms and their usage patterns, providing a robust framework for understanding how new lexical items are adopted and disseminated. Moreover, the exclusion of tweets without location data or those originating outside of Türkiye ensures that the findings are geographically relevant and accurately reflect regional linguistic trends.

In addition to its academic contributions, this study has practical implications for language policy and education in Türkiye. Understanding the newly emerged words can inform language planning and standardisation efforts, particularly regarding dictionary compilation and language teaching materials. By identifying which new words are gaining traction, policymakers and educators can develop more responsive and relevant language resources that reflect contemporary usage. This is particularly important in an era where digital literacy and online communication are becoming increasingly integral to everyday life.

In conclusion, the significance of this study on the geographical distribution of neologisms in Turkish cannot be overstated. It bridges a critical gap in the literature, offering new insights into the interplay between language, culture, and digital communication. The study provides a nuanced understanding of contemporary linguistic trends in Türkiye through its rigorous methodological approach and focus on real-time data. Its findings have far-reaching implications for academic research and practical applications, highlighting the dynamic and ever-evolving nature of language in the digital age.

RESEARCH QUESTIONS

1. What are the most frequently used neologisms by Turkish speakers on the social media platform X?
    1.1. How can the neologisms found on X be defined?
    1.2. Are the neologisms found in previous studies still used on X?
2. What are the categories of neologisms used on X?
3. Which categories are more prevalent as a way of neologism formation for Turkish people?

LIMITATIONS OF THE STUDY

The study's main limitation is data collection and analysis. As of 2024, Turkish social media text data publicly available to researchers is either very limited in sample size or completely private and cannot be accessed without authorisation.

Data collection from social media is expensive and time-consuming. In this study, more than 300,000 Tweets were gathered to analyse; however, with several years, appropriate equipment, and an extensive budget, millions of tweets can be gathered to analyse.

CHAPTER 1
LITERATURE REVIEW

The first chapter of this literature review will explain the history of neologisms and current practices among lexicologists and etymologists regarding this field. Neologism research on the Turkish language will be further explained. Finally, gathering data through social media platforms for linguistic research will be detailed.

1.1. BRIEF HISTORY OF THE STUDY OF NEOLOGISMS

According to Oxford Dictionary, neologism as a linguistic term was first used in the late 1700s and borrowed from the French word néologisme. The first formal and systematic studies of neologisms started as lexicography and etymology efforts. Various dictionaries catalogued neologisms. Examples of such works started appearing in English and French as early as the 17th century.

Although the neologisms have not yet been institutionalised or formally explained, there are dictionaries and essays covering the neologisms during the 17th, 18th, and 19th centuries. Even though it is hard to pin the first of its kind, various famous dictionaries or compilations date back to the period. One of the most famous examples of such dictionaries in English is *A New Dictionary of the Terms Ancient and Modern of the Canting Crew* (1698). The compiler of this dictionary is only known by his or her initials: B.E. The dictionary compiles canting crews of thieves, several gypsy tribes, cheats etc. Another example of such a dictionary is *Nouveau Dictionnaire des Mots Nouveaux* by Alfred Delvau (1874). It compiles new French words into one dictionary, giving it its name: A New Dictionary of New Words. However, no examples of Turkish dictionaries aim to compile neologisms.

American linguist Dwight Bolinger wrote the first formal and scientific compilation of neologisms as a periodic column in the scientific journal American Speech.

Bolinger's compilation was published between 1937 and 1940. However, the necessary technology to properly study neologisms developed after the 1970s, and both lexicographical and academic studies on neologisms gained pace from onwards. Academics such as Newmark (1988), Cabré (1993), and Rey (1974) defined neologism as a concept and created the outline of what constitutes a neologism. First, iterations of neologisms were based on Rey's definition, which states that neologisms are a new unit of lexical nature in a defined linguistic code. Indeed, this definition is still used today but improved upon. Newmark's definition in 1988 does not change Rey's framework but builds upon it. Newmark (1988) states that neologisms are "newly coined lexical units or existing lexical units that acquire new sense". This new definition builds upon Rey's framework and adds a semantic shift to the definition, as neologisms can be created through three different semantic processes. A semantic shift can give an entirely new meaning to an old word, narrow its meaning, or broaden its meaning. Karaağaç (2000) explains all four ways of semantic shifting for the Turkish language in his book *Türkçe'nin Dil Bilgisi* (p. 606). A semantic shift with a new meaning can be either semantic amelioration or semantic pejoration. "Yavuz" in Turkish used to have negative connotations, but it is currently the opposite, thus indicating a semantic amelioration. On the other hand, "canavar" used to mean "living being", whereas its new meaning now can be directly translated as "monster" and creates a semantic pejoration. Another type of shift is semantic broadening, which indicates that a word's meaning becomes more encompassing than its previous iteration. An example of such a word in Turkish would be the word "yurt", which shifted its meaning from a type of tent used as a home to "country". The final way of semantic shifting in Turkish is semantic narrowing, which indicates a word's meaning became less encompassing than its previous iteration. One example of semantic narrowing in Turkish is the meaning of "tünemek". It used to mean spending the night, but now it is exclusively used as a verb to describe the state of birds and other coop animals sleeping in their cage or coop.

Cabré (1993) defines four criteria to identify neologisms: date of appearance in the lexicon, exclusion from dictionaries, formal or semantic instability, and

speakers' perception of them as a novelty. Cabré further improves this definition by saying: “As objects of knowledge, neologisms are relative units that can only be identified when placed in a specific period, discursive context and enunciative perspective”. This new perspective presents a complete understanding of neologisms. Neologisms are novel words prone to formal or semantic changes, can only be identified in a specific context and time period, and do not appear in dictionaries as lexicologists. Adding a neologism to a general dictionary would mean the word is prevalent enough to warrant a dictionary entry. The dictionary criteria put forward by Cabré excludes specialised dictionaries such as neologism dictionaries, professional dictionaries, etc.

Social media and the Internet created a massive platform to facilitate the spread and usage of neologisms, and a shift to study neologisms on the internet occurred after 2010 (Tahiroğlu, 2014; Grieve et al., 2017; Pinto et al., 2020; Monderin, 2021). The mass amount of data with no observer complex facilitates compiling big corpora and doing more accurate frequency analysis. Frequency analysis for such data is conducted through the word’s occurrence frequency among the total number of words in the data.

1.2. CATEGORISING NEOLOGISMS

This chapter will further explain how neologisms are categorised. There are four different ways to categorise neologisms, namely function, origin, word formation, and source. It should be noted that the word formation category follows the word formation methods of the language neologisms that were created. These categories are further explained in the following sections to clarify the categories used in this study. These categories help understand the way neologisms were created and developed by speakers and are an essential part of the study.

1.2.1. Word Formation

There are various ways to categorise neologisms. One of the most common ways of categorising neologisms is through the word formation process behind its formation. Word formation processes are not universal rules; every language has its own set of processes for word formation. There are various studies analysing the word formation process of the Turkish language. According to Ergin (1994), the Turkish language has four processes for word formation that it uses to create new lexical units for new concepts. These four processes can be summarised as borrowing from a foreign language, compounding, reviving or compiling words, and creating new words. However, these categories are grouped broadly and do not explain the underlying details of the processes.

Karaağaç (2012) further details this word formation process and explains five methods of word formation in the Turkish language:

Affixation involves adding suffixes or prefixes to a root or stem to create new words with different meanings. This process modifies the base word to convey new ideas or functions. Conversely, inflexion forms new words or grammatical forms by changing the shape of words according to specific patterns and adjusting the word's form to express various grammatical features such as tense or number.

Reduplication is a less common method where new words are created by repeating or slightly altering part of the original word, as seen in the Turkish term "kapkara," which emphasises the meaning through repetition. Compounding creates new words by combining two or more base units. In Turkish, compounds like "hanımeli" (honeysuckle) demonstrate how merging separate elements forms new terms.

The auxiliary words method combines two prominent elements, one of which expands the meaning of the other. This approach is used in Turkish to form compound verbs, adpositions, and adverbial phrases, enriching the meaning and functionality of the words. The word order method utilises syntax, such as the placement of words in a sentence, to create meaning or emphasis. In Turkish, the element next to the verb often carries the sentence's focus, showcasing how word order can influence meaning.

Stress is another method for distinguishing word meanings or grammatical forms. In Turkish, stress generally does not change word meanings.

Eker (2013) states the word formation processes of the Turkish Language as follows:

Fixed Expressions (Kalıplaşma): This involves words or expressions becoming fixed in form and usage, deviating from their original grammatical roles. Examples include "Yaşar" (from "yaşa + ar") and "toptan" (from "top + tan").

Derivation (Türetme): New words are created by adding derivational affixes to base words or stems. Examples include "sarkaç" (from "sark+aç"), "birlik" (from "bir+lik"),

Compounding (Birleştirme): New concepts are formed by combining two words, often through compound phrases. Examples are "asbaşkan" (deputy chairman), "sağlık ocağı" (health clinic).

Blending (Karma): This involves creating new words by merging syllables or parts of two existing words. Examples include "albay" (from "alay+bay"), "arge" (from "araştırma+geliştirme"), and "gerzek" (from "geri+zeka").

Abbreviation (Kısaltma): Words are formed by combining the initials of longer phrases or words. Examples include "GAP" (from "Güneydoğu Anadolu Projesi") and "lab" (from "laboratuvar").

Revival (Derleme): Words that have fallen out of use in written language but are still used in spoken language are reintroduced into the written language. Examples include "alan" (field), "araç" (vehicle), and "asalak" (parasite).

Recovery (Tarama): Words that have been forgotten in written language but appear in old Turkish texts are revived. Examples include "bildiklü" (someone who knows many people), "bilecen" (someone who knows everything), and "bilegen" (someone who is knowledgeable).

Semantic Shift (Anlam Kayması): This involves a word acquiring a new meaning or concept, also known as metaphorical development.

Reverse Derivation (Ters (geri) türetme): This process involves deducing a word's derivational suffix from its form and using it to create related words. For example, reversing the suffix /l/ in "ayılmak" and "bayılmak" to create slang forms "ay-" and "bay-".

Coinage (Uydurma): New words are created intentionally without relying on existing morphological elements or rules. Examples include "uygar" and "bayan," which do not have scientific explanations for their creation.

Functional Change (İşlevsel değişim): The same form is used in different syntactic functions, such as the suffix (-sAl) used to form adjectives in "kumsal" and "uysal".

Borrowing (Ödünçleme): Words borrowed from other languages, such as "ahtapot" (octopus) and "körfez" (gulf) from Greek, "futbol" (football) and "egzoz" (exhaust) from English, and "derya" (sea) and "arzu" (desire) from Persian.

Kara (2011) coins the phono-semantic transformation (fono-semantik başkalaşma) for words created with phonetic changes. Kara examines the phono-semantic transformations under two main categories:

1. Phono-Semantic Transformation in Turkish Words: This includes transformations within and at the end of words between Turkish and Chuvash Turkish in the letter -r transforms to the letter -z and the letter -l transforms to the letter -ş, which he exemplifies with “bür” and “büz”.
2. Phono-Semantic Transformation in Borrowed Words: This involves changing borrowed words such as "abdal" to "aptal."

Yurtbaşı (2017) gives the following word formation methods for creating neologisms in the Turkish language:

1. Borrowing and loaning words: These types of neologisms are words that are borrowed or loaned from other languages into Turkish. For example “televizyon” in Turkish is a loan word from French, and fits into this category. The difference between borrowing and loan words is somewhat vague in Yurtbaşı’s definition. However, borrowings are not limited to words; they can also be phrases, letters, sounds, or a mode of speech. For example, “veni, vidi, vici” from Latin is a common borrowing for many languages, including Turkish.
2. Translation: Yurtbaşı states what is known as *calque* as neologisms created through translation. Calque words are created by literal translation of foreign words, either word by word or root by root. The most common example of calque in Turkish is “gökdelen”, translated from “skyscraper”.
3. Combining: This type of neologism is created by combining two words. One of the best examples in Turkish is “gizlilik sözleşmesi.”

4. Derivation: Derivation is the act of creating a new word by adding affixes or removing or changing morphological units. ”Geliştirici” is an example of such words in Turkish by Yurtbaşı.
5. Compounding: This type of neologism is created by compounding two words together. One of the most prominent examples in Turkish is “bilgisayar,” which is created by compounding “bilgi” and “sayar.”
6. Suffixation: These types of neologisms are created by adding a suffix to a root. Yurtbaşı chooses to categorise suffixation and derivation separately. It is probably one of the most dominant ways of word formation in the Turkish language. The example given for derivation, “geliştirici”, is also a neologism made by adding -tir and -ci suffixes to the root word “geliş”.
7. Clipping: These neologisms are created by removing certain parts of words. An example in Turkish would be “büt,” clipped from the word “bütünleme.”

However, word formation processes are not the only way of categorising neologisms. Neologisms can also be categorised according to their function, source, and coinage origin.

1.2.2. Different Ways of Categorizing Neologisms

Fang (2021) categorises neologisms into four standards for English:

1. Neologisms can be categorised through their functions. Referential neologisms fill a gap in a particular field to solve communication difficulties. For example, introducing “gizlilik ilkesi” to Turkish to fill a gap in communicating legality is a type of referential neologism. On the other hand, expressive neologisms are developed to introduce new forms to discourse. For example, “sunroof kız” to refer to a woman who leaves the front part of her headscarf open enough to partially show her hair is an expressive neologism.
2. Neologisms can be categorised through their coinage process. The coinage of a neologism can be divided into three sub-categories. It can be an old word

shifting its meaning (semantic shift). For example, albeit an old one, Karaağaç (2000) states that “oğlan” in Turkish used to refer to kids of any gender, whereas it now only refers to male children. Another way of coining a neologism is creating a new word to describe a new idea or process. For example, “e-okul” to refer electronic student database of elementary to high school students. Finally, a neologism can be a borrowed word from a different language. “Internet” or “televizyon” are examples of this type of coinage.

3. Neologisms can be categorised based on their word formation method. Fang's comprehensive categorisation includes derivations, compounds, phrases, shortenings (using initialisms, acronyms, clippings), semantic shift (broadening, narrowing, or completely changing the meaning), borrowings, and calques. These methods, though not all standard in Turkish, provide a rich understanding of neologisms in the English language.
4. Neologisms can be categorised according to their sources. Fang states the following sources for neologism categorisation: scientific words or phrases to describe new scientific concepts; political words or phrases to create political or rhetoric concepts; pop-culture words or phrases evolved from mass-media content; imported words or phrases originating from another language; trademark names turning into a reference for the products; nonce words used only for a specific and single occasion only, usually for a literary act; and inverted words, all of which are less prevalent in social media text.

1.3. NEOLOGISM RESEARCH ON THE TURKISH LANGUAGE

Various studies examine neologisms in Turkish throughout the years in studies such (Yurtbaşı, 2017; Tahiroğlu et al., 2014; Büyükkantarcığlu, 2000; Akyıldız, 2023; Dursun Önen, 2023). Moreover, there are books written on neologisms in Turkish, albeit lacking a scientific approach, presenting the observations of authors such as Lumpen Sözlüğü (Tülek, 2014). The studies on the subject are not supplemented with large quantitative data, and either present neologisms observed by the author or focus on specific neologisms contained in a limited environment such as a book or even sometimes limit themselves to one specific

neologism. One of the most critical studies on neologisms in Turkish with mass data is conducted by Tahiroğlu et al. (2014) to detect neologisms in online newspapers automatically. Tahiroğlu et al.'s program automatically scans every word in online newspapers and compiles them together. It then identifies the words that are not registered to TDK's online database. The words are grouped into an online database if they are frequently observed among the collected data. This database can be accessed only by TDK. TDK then manually selects the words and adds them to its dictionary. As of 2024, a portion of Tahiroğlu et al.'s neologisms such as "e-okul" (elektronik okul) and "e-devlet" (elektronik devlet" are added to the dictionary whereas words such as "gizlilik bildirgesi" are not given an entry on the dictionary. The program functions similarly to NeoTrack: Semiotomatic Neologism Detection (Janseen, 2005), which scans through the whole internet to detect neologisms by compiling the words not added to the dictionary and conducting frequency analysis per total number of words; it also requires manually editing some of the words out. According to Yurtbaşı (2017), the program used for Tahiroğlu et al.'s project is supplied by "The Global Language Monitor", which is a database company that analyses the language data on the internet to find current trends in the English language. The database produced by the project is not publicly available. However, its initial findings are presented within the project. If the initial findings are to be examined individually, it can be seen that some of the neologisms presented in the study are now fully included in TDK's online dictionary. The project's program can be furthered today as companies like Google have APIs allowing more accessible data collection from newspapers.

Another study conducted on a dataset collected from the internet is "The Language of Generation Z in the Axis of Generation Conflicts" (Çokol, 2020). Çokol identifies the meanings of neologisms on popular platforms such as Ekşi Sözlük and Uludağ Sözlük using slang and etymology dictionaries. Her study instead focuses on identifying the meanings of neologisms and explaining the reason behind their occurrences. Although the study does not include a frequency

analysis or any statistical data, it provides meanings for thirty-five different neologisms.

Another important study on Turkish neologisms is “Kırklareli merkez örnekleminde Z kuşağı gençlerinin sosyal medyadaki yeni kelimeleri kullanım alışkanlıkları üzerine nicel bir yaklaşım” (Şafak, & Bilginsoy 2020). The study groups native speakers according to generation theory, as the generation theory claims that individuals born in the same period are bound to have similar behaviours and characteristics. The study claims that the way Generation Z lives and sees life creates a need for neologisms. The researchers identified fifty neologisms and created a questionnaire with a five-point Likert scale. The scale goes from “I know the meaning and use the word frequently” to “I do not know the meaning and never use it”. The survey is done on high school children in Kırklareli. The students are given the questionnaire with a five-point Likert scale to identify how frequently they use the words selected for the study. The study identifies how the words were created simplistically, leaving out most of the neologism formats. Instead, it uses six general categories to identify them. These six categories are namely compound nouns, nouns, verbs, abbreviations, full borrowing, and partial borrowing. In contrast, these categories lay out general directions in which the word is created; it does not directly define the exact process of how it came to be. For example, an abbreviation does not indicate if the word is used via initialism, i.e. using the initial letters to create a word that should be read letter by letter, such as TDK, or acronyms, i.e. an abbreviation that can be read wholly, such as NATO The study analyses the frequency of how many of the words are created via these categories.

This study will aim to identify if the neologisms proposed by Tahiroğlu et al. (2017), Çokol (2020), and Şafak & Bilginsoy (2020) are still in use on social media as of 2023. Furthermore, it will further analyse the conclusions of previous studies in Turkey and whether they are used in social media today.

1.4. CURRENT NEOLOGISM RESEARCHES CONDUCTED THROUGH DATA GATHERED ONLINE

Apart from studies in Turkish, there are various studies in other languages cataloguing neologisms (Janseen, 2005; Smith, 2010; Levchenko, 2010). Similar to Tahiroğlu et al.'s project, they present a method to search through the web and analyse big data to compile candidate words for neologisms by comparing their findings to available dictionaries in their respective languages. However, these studies rely on newspapers and search engines. While they provide an extensive and steady chunk of data, they do not provide the natural real-time conversation data that can be gathered from social media. Unlike blogs or newspapers, social media provides a medium for real-time conversation for a diverse user base. Thus, the data that can be gathered from social media offers the possibility of finding neologisms, such as swear words or neologisms used by a small subculture, that may not appear on more formal mediums such as newspapers. Another crucial data that can be gathered from social media interactions is the metadata accompanying the conversation, such as emojis, geographical location, timestamps, pictures, etc. These may provide more insight and different approaches for researchers. APIs provided by the website can access data from social media websites. It offers diversity and size that any other written or oral data collection methods cannot provide.

Examining specific neologisms that have successfully entered mainstream usage provides practical insights into the process of lexical innovation. For example, the term 'selfie,' coined to describe a self-taken photograph, quickly became globally recognised and was added to dictionaries within a few years of its emergence. Similarly, the word 'blog,' a blend of 'web' and 'log,' has become a standard term in digital communication (Tagliamonte, 2016).

The success of these neologisms can be attributed to several factors, including their utility, memorability, and the social contexts in which they emerged. For instance, the term 'selfie' filled a lexical gap for a typical social media activity,

while 'blog' provided a succinct label for a new form of online publishing. These case studies illustrate how neologisms can quickly gain acceptance and become integral parts of the lexicon.

The spread of the term 'hashtag' provides another illustrative case. Initially used on Twitter to categorise topics, the term quickly spread to other social media platforms and even offline contexts. Research by Yang et al. (2012) demonstrates how the hashtag evolved from a simple tagging mechanism to a powerful social activism and marketing tool.

1.5. USING SOCIAL MEDIA TO GATHER DATA

Shoemark (2020) proposes three different methods of gathering data from X and identifies the pros and cons of each method. Three methods presented in Shoemark’s work are “Streaming API” with sampling endpoint and filtering endpoint and “Search” API. While Search API can retrieve historical data for up to one week, it requires a more manual data collection process from the researcher compared to Streaming API methods. Streaming APIs are developer access to Twitter's Streaming API, which is a powerful tool that provides developers with real-time access to Twitter data. This API allows users to collect tweets as they are posted, offering a continuous stream of public tweets based on specific criteria set by the user. These criteria include keywords, phrases, hashtags, or user accounts.

The Streaming API is particularly valuable for researchers and developers who require up-to-the-minute data for analysis, monitoring, or application integration. It enables tracking trending topics, monitoring public sentiment, and gathering data for machine learning projects.

There are several endpoints within Twitter’s Streaming API, each designed for different use cases:

1. Filter Endpoint: This allows users to filter the real-time stream of tweets based on keywords, user IDs, locations, and other parameters.
2. Sample Endpoint: Provides a small random sample of all public tweets.

The data provided by the Streaming API includes not just the text of the tweets but also metadata such as user information, tweet timestamp, geolocation data, and more. This rich dataset is instrumental for detailed analysis and understanding of social media dynamics.

To use the Streaming API, developers must authenticate their requests via OAuth, ensuring secure and controlled access to Twitter's data. The API delivers data in JSON format, which is easy to parse and manipulate using various programming languages and tools.

Twitter's Streaming API is a critical resource for anyone needing real-time insights and data from the vast stream of conversations on Twitter. It leverages social media's immediacy and reach to facilitate various research, analytics, and real-time monitoring applications.

For the purpose of this study, a sample endpoint will be used to filter out the data without GeoTags. GeoTags are the location markers on Tweets, indicating the location of the user's real-life location. Non-GeoTagged Tweets will present null data for the purposes of this study. Therefore, the sample endpoint provides a substantial way to eliminate such data from the study.

Grieve et al. introduce a method for mapping lexical innovation on American social media, using a multi-billion-word corpus of Tweets collected between 2013 and 2014. The researchers extracted 54 emerging words from the corpus by searching for words that were very uncommon at the end of 2013 but whose use rose dramatically throughout 2014. They then map the origin and spread of each of these words. Based on these results, they identify five main regional patterns

of lexical innovation and emerging words on American Twitter, primarily associated with the West Coast, the Northeast, the Mid-Atlantic, the Deep South, and the Gulf Coast. Grieve first compiles the Tweets with GeoTags from the US and dissects each Tweet word by word. The words are then eliminated if they occur less than 500 times in the corpus. The remaining words are then eliminated if they are proper nouns to eliminate, including brands or products that got popular recently from the list. Next, they eliminate every word with an entry in Merriam-Webster's *Dictionary* to focus the analysis on relatively new word formations. They further eliminate the abbreviations and initials primarily related to the medical industry since their frequency on Twitter is increased through job postings due to Geo-Tagged employment advertisements. The remaining words are compiled through a list of eighty-one new emerging words. Some of the examples from the study are “cosplay” (costume role play), “waifu” (wife), and bruuh (bro). These words are then grouped together if they have alternative spellings since the different word forms, in this case, do not change the meaning. New emerging words such as “bruh” in the study have variations that account for some of the 83 words found in the study. For example, “bruh” has ten different variations with various numbers of increased -u and -h letters, such as “bruuhh” or “bruhhhhhh”. The study finds thirty-eight new emerging words when all the variations are accounted for as one.

1.5.1. Digital Communication and Language Change

The advent of digital communication has introduced new dynamics in language change. Social media platforms, instant messaging, and other forms of digital interaction create new spaces for linguistic innovation. Tagliamonte (2016) highlights how digital communication accelerates language change, particularly among younger generations, who are often at the forefront of adopting and disseminating new terms.

Tagliamonte’s research on digital communication underscores the pivotal role of social media platforms like X and Facebook in the swift dissemination of

neologisms. By enabling users to coin and circulate new terms in real time, these platforms foster their widespread adoption and standardisation. Tagliamonte posits that digital communication engenders a more dynamic and fluid linguistic milieu, where language evolution can manifest more swiftly and conspicuously than in conventional, face-to-face communication.

Crystal's (2006) comprehensive exploration of the internet's influence on language use is a testament to the profound impact of digital communication. His study reveals how the internet has revolutionised language use, introducing a diverse range of new genres and modes of communication. Crystal argues that by encouraging linguistic creativity and the formation of new lexical items, Crystal argues that digital communication prompts users to adapt language to the unique constraints and affordances of new media.

Socio-cultural factors such as economic changes, migration, and cultural exchanges profoundly influence language formation. Neologisms often mirror these dynamics, serving as barometers of broader social trends. Milroy (2002) delves into how social networks and communities of practice contribute to linguistic innovation, underscoring the role of interpersonal interactions and cultural exchanges in shaping language.

Milroy’s social network concept highlights the importance of social relationships in language use and change. She argues that individuals’ linguistic practices are influenced by their social ties, with dense and multiplex networks promoting linguistic stability, while loose and uniplex networks encourage linguistic innovation. This framework helps explain why certain neologisms gain traction in specific social groups or regions, reflecting the social dynamics at play.

Milroy and Milroy (1985) further elaborate on the role of social networks in language change, suggesting that tightly-knit communities with strong social ties tend to resist linguistic change, while more loosely connected communities are more open to adopting new linguistic forms. This dichotomy provides a useful

lens for understanding how neologisms spread through different social and cultural contexts.

1.6 OVERVIEW OF X AS A LINGUISTIC PLATFORM

X serves as a microcosm of contemporary language use, providing real-time data on linguistic trends. Its structure, which limits posts to short messages, encourages the creation of succinct and innovative expressions, making it an ideal platform for studying neologisms. Dannet and Herring (2007) highlight how X facilitates the rapid spread of new terms by enabling users to interact without geographical and temporal constraints.

The character limit on X forces users to be concise and creative, often leading to the invention of new abbreviations, acronyms, and slang terms. The platform's real-time nature allows for the immediate sharing and dissemination of neologisms, creating a dynamic environment for linguistic innovation. X's global reach also ensures that new terms can quickly spread across different linguistic and cultural contexts.

Rheingold (2000) discusses the concept of virtual communities, highlighting how online platforms like X facilitate the formation of linguistic communities that transcend geographic boundaries. These virtual communities play a crucial role in creating and spreading neologisms as users adopt and propagate new terms within their networks.

Digital platforms facilitate the rapid spread of neologisms by connecting users across vast geographical areas. This interconnectedness allows new terms to gain traction quickly as they are shared, retweeted, and adapted by a global audience. The virality of neologisms on platforms like X exemplifies the role of digital communication in linguistic innovation (Crystal, 2011).

Research by Zappavigna (2011) demonstrates how hashtags on X can drive the spread of new words and phrases. Hashtags serve as aggregators for content related to specific topics, allowing users to participate in global conversations and contribute to the propagation of neologisms. The study highlights the role of digital platforms in creating and sustaining linguistic trends, emphasising the importance of social media in contemporary language change.

Similarly, Mace (2013) examines the role of memes in spreading neologisms on social media. Memes, which are often humorous or satirical images with text, frequently introduce new words and phrases that quickly become part of the digital lexicon. Mace argues that memes' visual and viral nature makes them powerful tools for linguistic innovation.

Case studies of specific neologisms that gained popularity on X provide insights into the mechanisms of digital linguistic innovation. For example, the spread of terms like 'tweetstorm' (a series of connected tweets) or 'hashtag' (a keyword prefixed by a # symbol) illustrates how new words emerge and proliferate on social media. These case studies highlight the role of digital platforms in shaping contemporary language use.

The case of 'tweetstorm' shows how new terms can quickly become part of the digital lexicon. Coined to describe a series of connected tweets posted in quick succession, the term reflects both the platform’s technical affordances and the communicative practices of its users. Similarly, 'hashtag' has become a ubiquitous term for tagging and categorising content, demonstrating the influence of digital platforms on language.

Herring et al. (2013) explore the evolution of internet slang, focusing on the spread of neologisms across different online communities. Their study highlights the role of social media in facilitating linguistic convergence as users adopt common terms to participate in online conversations. This process of linguistic

alignment contributes to the standardisation of neologisms within the digital lexicon.

Using APIs to collect data from platforms like X allows researchers to access large datasets, capturing a wide range of linguistic phenomena. This method is particularly effective for studying the frequency and distribution of neologisms (Bamman et al., 2014). The data collected can be analysed to identify patterns and trends using new terms, providing valuable insights into how neologisms spread and become integrated into everyday language.

APIs provide researchers with real-time access to data, enabling the collection of large-scale datasets that reflect current linguistic trends. This method allows for analysing temporal patterns, such as how quickly neologisms gain popularity and whether their usage is sustained over time. Additionally, APIs enable the collection of metadata, such as user demographics and geographic locations, which can be used to explore linguistic innovation's social and spatial dimensions.

CHAPTER 2
METHODOLOGY

2.1. RESEARCH DESIGN

The study employs quantitive analysis to determine possible neologisms. It mainly follows two practical frameworks, one for data collection and one for data analysis. This chapter will lay out their applications to this study. Shoemark's (2013) data collection method for Discovering and Analysing lexical variation in social media text is laid out.is used as a framework to gather data through X. Social media website X is used as a source of data, and a corpus is compiled by using X API to access Tweets and user bios, a short section in a user's profile that they use to introduce themselves. The data is collected through the Streaming API of X. This framework is further explained in section 1.5.

The data is analysed using the framework laid out by Grieve et al. (2017) in the study *Mapping Lexical Innovation on American Social Media*. Grieve et al. lays out a framework for relevant frequencies, which are adjusted to account for the amount of data difference between the corpus and how to account for unrelated variables. The study is explained in further detail in chapter 1.5. One minor adjustment to Grieve et al.'s framework is that this study's analysis will not require a normalisation adjustment for tweets per day since Grieve et al. use this on a corpus compiled through a filter endpoint, which is not available as a means of data collection for this thesis.

The study's initial step requires data gathering to create a corpus, which is then used for the analysis since a comprehensive social media corpus in Turkish is not available for this purpose.

2.2. DATA COLLECTION

The first step of the study included compiling a corpus to analyse since corpora in the Turkish language do not meet the requirements of the study. *TrTenTen: Corpus of Turkish Web developed by Sketch Engine company and Masaryk University offers corpora consisting* of more than ten billion words. However, the corpora are gathered through websites using web crawlers and do not gather data through social media websites as it is against their terms and conditions. Çöltekin (2020) also compiled a corpus on Twitter, *A Corpus of Turkish Offensive Language on Social Media. However, this corpus only compiles 36,232 tweets from* 2018 to 2019. This corpus is outdated for the study of neologisms and does not provide large enough data to work on. Thus, compiling a new corpus for the analysis is a necessary step.

This section outlines the methodology employed to collect and clean tweet data originating from Türkiye in 2023. The data collection process was executed using Tweepy, a Python library that allows easy access to the X API. The subsequent data cleaning process involved using the Pandas library, which is commonly used for data manipulation and analysis.

2.2.1. Data Source

The primary data source for this study is X, a social media platform known for its widespread use and real-time dissemination of information. X provides a unique opportunity to collect large volumes of user-generated content, which is ideal for tracking the emergence and spread of neologisms. Social media data proves further useful in eliminating observer paradox as the interactions between users occur naturally without facilitation by a researcher with methods such as questionnaires or certain topics to converse on, and it also eliminates the sense of being recorded to be studied. Thus, the interactions between users are far more organic than interactions specifically recorded to study with the participants' knowledge.

X API provides access to necessary data for the study, allowing specific parameters to be set to collect Tweets based on location, language, time frame, and more. Furthermore, X API allows users to access a large amount of data at once, eliminating the need to manually collect each Tweet and speeding up the process considerably. For the purposes of this study, especially in setting criteria for a specific time period, language and GeoTag were important so as not to bloat the corpus with irrelevant Tweet entries.

2.2.2. API Access and Authentication

In this study, the X API's sample endpoint was used. This endpoint allows for real-time tracking of Tweets containing specific information. GeoTag data and the language parameter were set within the sample endpoint to gather Tweets only from Turkey. The focus on tweets from 01/01/2023 to 31/12/2023 ensured that the dataset was current and relevant to the period of interest.

The following criteria were established for data collection:

Location: Turkey
Time Period: 01/01/2023 – 31/12/2023
Language: Turkish
Sampling process: Random

By setting these parameters, the study collected only Tweets relevant to the research context. The collection process was conducted over a period of time, capturing a wide range of tweets from different users across Turkey.

Authentication via OAuth (Open Authorization) was required to access X API. This protocol allows third-party applications to interact with X's servers securely. To access X API, an X developer account is necessary. Researchers use this account to access data within the perimeters allowed by X. The developer

account generates unique API keys, API secrets, access tokens, and access secrets unique to each of its users, which in turn can be used by third-party libraries to interact with the API if the users choose to do so. Alternatively, X API provides a user interface that can be utilised to access its facilities but is rather limited. X API fully supports third-party libraries and shares a number of them created for specific purposes on their developer platform. All the libraries enforced by X API are open source and can be accessed by anyone regardless of whether they can authenticate to use the X API. However, accessing libraries without authentication will result in an error when accessing API endpoints, and the users will be unable to utilise the libraries.

2.2.3. Data Collection Through X API

The Tweets collected for the corpus are compiled through four parameters. The data collection method opts for a sample endpoint rather than a filter endpoint as the filter endpoint requires means not available for this study. For sampling, the date of the Tweets, geo-location of the Tweets, language of the Tweets, and selection method of the Tweets were set up as necessary parameters. Tweets chosen otherwise are entirely randomised. The Tweets were randomly selected from the dates 01/01/2023 to 31/12/2023, filtered to exclude any language other than Turkish, and excluded any Tweet shared from a place outside of Turkey.

This process is done using Python and a third-party library called *Tweepy*. Tweepy is the most popular X API library, allowing users to access the API and X endpoints easily. It is an entirely free and open-source library. It is selected to eliminate the need to manually adjust the code to compile data each time the platform is being scanned. Typically, using X API without a library means the researcher must handle many crucial details such as HTTP requests, authorisation, rate limiting, and serialisation; Tweepy is not specialised in compiling language data, etc. However, *Tweepy* automatically handles these sections. This is rather important as X API limits the number of queries and requests from the API to execute tasks on a timely basis. A developer account

may not request more than one thousand queries per fifteen minutes, accessing one thousand Tweets per fifteen minutes. These one thousand Tweets also include the Tweets that do not meet your filtering criteria since the compilation method is a random sampling; it should be noted that even though the library is famous for compiling linguistic data, it is not explicitly written for the purpose. The library allows users to access the endpoints and execute commands on X API as they usually can, such as sending Tweets, searching for trends, gathering Tweets, planning a posting schedule, etc. However, the user still needs to write prompts in Python to effectively utilise the library according to their needs, as the library does not provide a shortcut of commands. It is an enabler rather than a tool that operates independently.

2.2.4. Clearing the Data

The compiled data for the corpus includes raw information irrelevant to the study. The compiled data must be cleared first to facilitate a more straightforward analysis by eliminating unrelated data such as emojis and to prevent the data from being skewed by duplicates, user names, and Retweets. Data cleaning is handled through Python using Pandas, a free, open-access library.

Pandas offers data structures and tools to clean, normalise, visualise, inspect, and save data. It is mainly used for data sciences and AI learning. The study uses the library to clean and save the data gathered through the API. *Pandas* is used to clean the data first. This step requires a few queries with the library before saving the data. The first step is to clear any duplicates in the data. Since the query limitations per fifteen minutes of the X API limit the users from gathering mass amounts of data in one scraping session, the process requires multiple data collection sessions. Thus, some Tweets are unintentionally saved more than once in the raw data. To solve this, *Pandas* is used to scan through Tweet IDs. Each Tweet has a unique Tweet ID only visible through the API. However, these IDs prove helpful in eliminating duplicated data. The library is used to can all Tweet IDs and eliminate duplicate Tweets sharing the same Tweet ID. Next, the

library is used to eliminate non-textual content within the data. This process also involves the elimination of URLs and pictures as they are not required for the study. This requires identifying what constitutes a URL or non-textual content and writing codes accordingly to prevent the accidental deletion of crucial textual information. The final step of data cleaning was to filter Retweets out of the data. On the X platform, users can share Tweets posted by other users. This is called Retweeting. Retweets may include an original text alongside the shared text, but it is optional. Thus, Retweets without an accompanying original text are cleaned from the data to prevent hot trends or Retweets of famous users from skewing the data. The raw data enables this by tagging Retweets with the initials of *RT* before them. Therefore, the *Pandas* can be used to filter out any data containing this specific tag.

The cleaned data is then organised in columns using the same library. Each column gives the necessary information for the study, namely Tweet ID, username, Tweet itself, and geographic location. These saved files are manually filtered to remove Tweets not shared from Turkey. This process is handled manually rather than automatically by using a library because the geographic information of Tweets is based on the information given by the user, but not every piece of information was in the Turkish language or contained either the words “Turkey” or “Türkiye” in it. Some geographical information only states the name of a city, a neighbourhood, or a village. In addition, some of the geographic information was written by users in alphabets or fonts that would give out false clearings for the library. For example, a user with geographic information written as “Türkiyé” instead of “Türkiye” would be cleared out by the library. However, the study accounts for these geographical locations as valid values.

2.3. DATA ANALYSIS

The data analysis for the study requires the textual parts of the Tweets to be segmented and counted word by word before analysing any numeric value. Grieves et al. (2017) handle this by compiling each word separately. However,

this proves to be problematic for an agglutinative language like Turkish. Agglunating languages string together morphemes to create lexical units, and each morpheme corresponds to one syntactic feature. Turkish mainly uses suffixes for inflexion. The Turkish language inflates a word with a number of suffixes without changing the meaning of the root. Therefore, the data may be skewed heavily if this is not accounted for in the analysis, as two data entries containing the same lexical unit with different inflexions will be detected as different word groups. For instance, “mükemmel” (perfect) is a word with no inflexions and would be counted as a unique lexical unit. On the other hand, “mükemmeldi” (it was perfect) is the same word with the suffix -di, which only adds tense information and does not change the word’s original meaning. However, it would also be counted as its unique lexical unit.

2.3.1. Tokenization and Morphological Analysis

To overcome the problems mentioned in section 2.3, the words must first be analysed in a morphological analyser tailored explicitly to the Turkish language. Çöltekin (2014) created an open-source and free morphological analyser called TrMorph, tailored explicitly for Turkish. *TrMorph* does not automatically tokenise sentences, but it morphologically analyses any lexical unit and separates it into morphemes. This was utilised by first tokenising the textual data and then running it through *TrMorph*. The tokenisation was handled through a Python library called *TRNLP* (Tr Natural Language Processing). *TRNLP* provides not only a tokenisation library but also a morphological analyser. However, *TrMorph* proves to be more accurate based on a test of one hundred words conducted for this study to compare their accuracy. Therefore, *TRNLP* is only used to token each word without analysing the morphemes.

These tokens are then manually analysed to see if there are a string of words rather than singular words that may need to be analysed separately as neither of the current language models has an automated way of controlling this. The identified word strings are compiled separately to be analysed as a whole string

rather than separate units. For example, “sunroof kız” is one of the identified strings analysed separately from singular units. That is to not skew the data with instances when “sunroof” or “kız” are used separately. The tokenisation process is kept relatively simple; the end product of a tokenised sentence yields a result ready to be morphologically analysed, but it does not further the analysis. However, it should be noted that *TRNLP* successfully tokenises punctuations, and the punctuation tokens were not used for the morphological analysis. So a sentence run through the tokeniser yields a result such as:

The original sentence: lütfen cahilliğimi mazur görün sunroof kız ne demek?

Tokenized sentence: [‘lütfen’, ‘cahilliğimi’, ‘mazur’, ‘görün’, ‘sunroof’, ‘kız’, ‘ne’, ‘demek’, ‘?’]

After creating the tokens, each token was analysed through *TrMorph* to determine inflected variations of the same word units. Inflected variations are counted as one group rather than unique groups. *TrMorph* identifies both the root and morphemes attached to it as long as the root word is included in the TDK’s online dictionary. Suppose the root is not in the dictionary. In that case, it gives a null result, which is also helpful for the purposes of this study as neologisms are not included in dictionaries, as previously stated. However, word strings and semantic shifts may appear as valid data. Null results were manually analysed to ensure the results. A tokenised word analysed through *TrMorph* gives a result such as the following:

Tokenized unit: cahilliğimi

Morphologically analyzed unit: cahil<Adj><lik><N><p1s><acc>

The abbreviations on the analysed words directly state the morpheme added to the root. The example given above shows that the word’s root is “cahil”, and it is

an adjective as signified by “<Adj>’. The following marker, “<lik>”, signifies that it is a derivational morpheme, as the analyser leaves derivational morphemes as they are. The following “<N>” indicates that “cahillik” is a noun. So it concludes that the adjective “cahil” is turned into a noun by getting the derrivational morpheme -lik The following marker, “<p1s>, indicates that the word is in first person singular form. Finally, the last marker, “<acc>”, signifies that the word is accusative. Each morphologically analysed word with the same root is then manually analysed to see if they only have inflexions. If they only have inflexions, they are grouped as the same word unit rather than separate units. Any findings with the same root but with a derivational morpheme are accepted as separate word units. This is called lemmatisation, and Grieve et al. (2017) also put forward that it is a possible step for researchers to conduct if the analyst chooses to do so.

2.3.2. Frequency Analysis

The frequency analysis was conducted using the framework of Grieves et al. (2017) in their work Analyzing lexical emergence in Modern American English Online. The framework first normalises the frequency of each word day by day. This allows for further analysis using the Spearman correlation coefficient to identify monotonic patterns. However, Grieves et al. state, “Frequencies were normalised PBW to allow for results to be expressed in whole numbers, as this analysis is focusing on sporadic forms; normalising by PBW does not affect the results of the analysis.” Since the data collection method is different for this study, it is impossible to identify the same monotonic patterns and make a day-by-day comparison. Unlike Grieves et al., this study uses a sample endpoint rather than a filter endpoint. The filter endpoint allows the researcher to gather day-to-day data, whereas the sample endpoint allows the researcher to compile randomly selected data. Thus, the sample endpoint provides unreliable results for such analysis.

The primary frequency analysis method used for this study is laid out by the same study as the average frequency per million words. This method requires a word to be used at least once per one million words (PBW). However, the researchers state that the number can be adjusted based on available data. Since the framework is built upon a corpus that compiled data for one and a half years with millions of entries and over 8.9 billion words, it has significantly larger data to analyse. On the other hand, this study was conducted through 327.262 entries with a total word count of 2.463.075. Therefore, the threshold should be adjusted accordingly. If the threshold is adjusted directly based on the size of both corpora rounded up, the threshold is set to once per every 250,000 words.

The words or word strings identified with the analysis will then be eliminated if they occurred less than five times in the corpus. This threshold is adjusted from Grieve. Et al.'s framework eliminates anomalies, as they also eliminate words that occurred less than 500 times in the whole corpus. If this study's threshold were adjusted directly based on the size difference of corpora, the threshold for this study would be at least 0.3 occurrences in the whole study. Any word in the corpus would be a potential neologism candidate. However, a minimum number is decided with the purpose of eliminating accidental spelling errors, as a consciously made spelling variant would be more likely to occur more than once. The remaining words are filtered through TDK's official dictionary to detect whether they have dictionary entries or not, and the results are filtered manually to exclude relatively old words that do not have dictionary entries yet. For example, "Anadolu Lisesi" is a word string not found in the TDK's dictionary and identified by this study. It is also a neologism suggested by Tahiroğlu et al. (2014). However, the word string has been used by Turkish speakers in both formal and informal settings for decades, as the word string has been in use since 1976, following a circular from the Ministry of Education. Neologisms for this study are relatively new forms of expressions entering into the general usage on social media website X. For the purpose of this study, word forms are defined as case-insensitive strings of alphabetic characters, hyphens, and apostrophes. Creative spellings and acronyms are also included in this definition. They combine

quantitative and qualitative approaches to comprehensively analyse the geographical distribution and usage patterns of neologisms in Turkish. This design allows for a robust analysis of large-scale data while providing the contextual depth necessary for understanding the sociolinguistic dynamics underlying neologism adoption. Data will be collected through X Api via Python and cleared with Pandas before analysis.

2.4. ETHICAL CONSIDERATIONS

Gathering data through social media is subjected to ethical limitations, just as every other data collection method. This section will give further information regarding the ethical limitations of data collection through social media and how they are implemented for this study.

2.4.1. Informed Consent

Given that the data is collected from public Tweets, obtaining informed consent from individual users is not feasible. However, ethical guidelines for social media research emphasise the importance of respecting user privacy and data protection. Only publicly available data is collected for the study, and any identifying information will be anonymised to protect user identities. Any data usually present in private user profiles (i.e. data that is only open to be viewed by people that the user allows) is not collected in any way or format. Gathering data through publicly available sources does not require personal consent from any involved party. Social media terms and conditions also state this and give a way of taking their consent back to users by simply setting their profile settings to private. This setting completely blocks access to a user’s profile through the API or regular means. However, if the user comments under a publicly visible Tweet, their comment will be accessible via the API and other platform users.

2.4.2. Data Anonymization

To ensure the confidentiality of users, all personal identifiers (e.g., usernames and profile pictures) will be removed or anonymised in the dataset. This will prevent the identification of individual users and mitigate privacy risks.

2.4.3. Compliance with Platform Policies

The study will adhere to terms of service and data usage policies. This includes ensuring that data is used solely for academic research and that the corpus cannot be shared with the public. Any shared data type can only be in the form of Tweet IDs.

CHAPTER 3
FINDINGS AND DISCUSSION

This chapter presents and discusses the study's findings. In the first section, neologisms identified in the corpus are laid out, and their word formations are explained. The second section defines the meanings of the neologisms identified in this study.

3.1. DATA COLLECTION RESULTS

A total of 327.262 Tweets with a total word count of 2.463.075 were compiled for the corpus. The findings were then grouped into lexical units as singular words or word strings (i.e. neologisms created by combining more than one word). Lemminization and semantic shifts were also taken into consideration for the analysis. The neologisms that occur once per 250,000 words are identified as possible candidates. The data suggested 74 neologisms in the Turkish language. Some of which overlap with neologisms proposed in previous studies.

3.2. IDENTIFIED NEOLOGISMS

The table below lists the neologisms with significant frequencies in the analysed data. Seventy-four unique entries were identified as neologisms through the analysis. Twenty-one of those neologisms were identified by previous studies (Tahiroğlu et al., 2017; Çokol, 2020; Safa & Bilginsoy, 2020). Fifty-two neologisms emerged from the data collected for the study.

Table 1. *Collected Data*

	Neologisms	**Frequency per 250,00 words**
1	Afk	2.35
2	Akmak	1.63
3	Alfa	1.22
4	Anlık	1.43
5	Aşko	5.71
6	Atar yapmak	1.12
7	Banlamak	3.98
8	Ben şok	1.84
9	Boomer	1.02
10	Boş yapmak	3.67
11	Buga girmek	2.14
12	Cringe	3.57
13	Çar	1.12
14	Dişil enerji	1.33
15	DM	6.94
16	Düşmek	3.47
17	Efso	2.24
18	Engel atmak	4.08
19	E-reçete	1.02
20	Eril enerji	1.22
21	Erko	2.24
22	Fake hesap	2.04
23	Favlamak	4.90
24	Füze atsaydın	1.43
25	Ghostlamak	4.39
26	Gizlilik politikası	1.02
27	Glow up	1.22
28	GOAT	1.84
29	Gümlemek	1.53
30	Güno	5.82
31	Halis mi?	3.16
32	Influencer	2.55
33	Kanzi	7.24
34	Konum atmak	1.43
35	Köpke	1.73
36	Manifestlemek	3.67
37	Meme	6.33
38	Mezuna kalmak	4.80
39	Müko	5.41
40	Ne münsaebo	1.02
41	NPC	1.63
42	Patlamak	1.43
43	PC	6.22
44	Pick me	3.16
45	PP	4.80
46	R yapmak	3.06
47	Reyiz	2.14
48	Roket atsaydın	1.53
49	Salmak	1.43
50	Shiplemek	3.78

Table 1. *(Continues)*

	Neologisms	Frequency per 250,00 words
51	Sıfadül eşgal	1.22
52	Sigma	1.12
53	Slay	1.94
54	Stalklamak	4.90
55	Story	2.76
56	Sunroof kız	2.14
57	Takipçi kasmak	4.18
58	Tilt olmak	2.65
59	Triggerlanmak	1.94
60	Vibe	3.57
61	Yargı dağıtmak	1.84
62	Yeto	1.73
63	Yıkık	2.35
64	Youtuber	3.78
65	Yükselmek	1.63
66	Yürümek	2.04
67	Zırvana	1.12

3.2.1. Neologisms Overlapping with Previous Findings

The data puts forward twenty-one neologisms identified by the previous studies. The data shows a statistically significant frequency of the three neologisms Tahiroğlu et al. proposed (2017). The remaining 17 previously suggested neologisms are divided between Çokol (2020) and Safa & Bilginsoy (2020), with eight neologisms for Çokol and nine neologisms for Safa & Bilginsoy. The table below lists the relevant neologisms and the researchers that first proposed them.

Table 2. *List of Previous Findings*

No	Neologism	Researchers
1	E-reçete	Tahiroğlu et. al. (2017)
2	Pc	Tahiroğlu et. al. (2017)
3	Gizlilik politikası	Tahiroğlu et. al. (2017)
4	Akmak	Çokol (2020)
5	Atar yapmak	Çokol (2020)
6	Yükselmek	Çokol (2020)
7	Yürümek	Çokol (2020)
8	Boş yapmak	Çokol (2020)
9	Stolklamak	Çokol (2020)
10	Yargı dağıtmak	Çokol (2020)
11	Mezuna kalmak	Çokol (2020)
12	Stalklamak	Şafak & Bilginsoy (2020)

Table 2. *(Continues)*

No	Neologism	Researchers
13	Youtuber	Şafak & Bilginsoy (2020)
14	PP	Şafak & Bilginsoy (2020)
15	Banlamak	Şafak & Bilginsoy (2020)
16	Konum atmak	Şafak & Bilginsoy (2020)
17	Story	Şafak & Bilginsoy (2020)
18	DM	Şafak & Bilginsoy (2020)
19	Favlamak	Şafak & Bilginsoy (2020)
20	Engel atmak	Şafak & Bilginsoy (2020)

In addition to neologisms in *Table 2*, the analysis showed a statistically significant frequency for some neologisms proposed by Tahiroğlu et al. These neologisms can be listed as: “Anadolu Lisesi” with a frequency of 2.65, “Doğalgaz” with a frequency of 1.94, “Yargıtay” with a frequency of 1.02, and “Açıköğretim” with a frequency of 2.04. However, these neologisms are excluded from the study as they do not fit this study’s criteria for neologisms.

“Anadolu Lisesi” is a word used to describe high schools in Turkey that aim to educate bright students with the aim of high academic success. “Anadolu Lisesi” was established in 1975 following the Ministry of Education circular. They were previously known as “Maarif Kolejleri.” However, following a mandatory name change from “kolej” to “college” by law, the new name was adapted to describe such high schools. The term has been in formal use at least since 1975. Cabré (1993) states that “as objects of knowledge, neologisms are relative units that can only be identified when placed in a specific time period, discursive context and enunciative perspective.” Indeed, “Anadolu Lisesi” was a neologism since it is a coined term and fits every criterion to be a neologism, but only when examined through place in the specific time period that it was coined, namely in 1975. The term not having a dictionary entry may have various reasons, but it is hard to give a definitive reason since TDK’s criteria for new dictionary entries are somewhat vague.

“Yargıtay” (Court of Cassation) falls short of fitting the neologism criteria of this study for the same reason as “Anadolu Lisesi”. Although the etymology of

“Yargıtay” is unclear, it may be a word created through a word-formation process called blending. Blending is a process in which parts of words are combined together to create new words. In this case, “Yargıtay” may be created by blending “yargı” with “kurultay” by clipping the head of “kurultay” and adding it to “yargı”. This is further supported by the fact that -today is not a derivational affix used in Turkish, thus making it more likely to be a blended word. According to Yargıtay’s (Court of Cassation) own website, the name was adopted on 10.01.1945 and replaced “temyiz mahkemesi”. Therefore, it also does not meet the recency criteria.

“Açıköğretim” is a type of distance learning method for the university level. According to Anadolu Universty’s website, which established the first faculty for it in Turkey, the faculty was established in 1984. Therefore, it also does not meet the recency criteria. However, there is a different angle to examine here. Although “açıköğretim” is not in the dictionaries, “açık öğretim” is in the TDK’s dictionary. It is unclear whether the term was supposed to be written as compounded at first, and the rule later got changed by TDK, or whether the word was always supposed to be written correctly. The name “açıköğretim” is an overlook from Anadolu University. It is unclear whether TDK changed the ruling of compound words or a specific ruling for this term. TDK does not have a dedicated source to track the changes and additions to rules and dictionaries. Dictionaries published by TDK between 1980 and 2017 are examined from its archives to see whether the term had any changes, but they yielded no results. The first entry for the term is seen in later dictionaries published in the 2010s, and they state the term as “açık öğretim” starting from the first entry. The Wayback Machine is also utilised to determine if TDK changed the grammatical rules of compound words. The Wayback Machine is a publicly available and free tool that archives snippets of web pages. It can be used to examine a web page as it was in a given day, month, or year as long as it has a snippet archived from the desired time. However, this effort also yielded no results as any information stored on The Wayback Machine does not indicate a rule change that may cause this.

“Doğalgaz” (natural gas) is the final word with a statistically significant frequency. “Doğalgaz” is a word that is borrowed through calque to Turkish. It has been in use at least since the 1980s, as it is the decade that natural gas started to be imported to Turkey. However, there is a challenging issue with the word that is hard to resolve. According to TDK, the word is written as “doğal gaz”. There are no changes on compounding rules or dictionary entries that indicate “doğal gaz” was supposed to be written as “doğalgaz”. So, it may fit to be a creative spelling or a common mistake that may eventually replace the proper spelling. However, there may be disambiguation surrounding the word, and the different spellings may be just a clash of different authorities on the subject rather than a conscious or unconscious way of altering the spelling. There are non-TDK operated dictionaries, especially specialised dictionaries on sciences, such as *Türkçe Bilim Terimleri Sözlüğü* that state the term should be written as adjointed. This clash seems to even persist through official institutions. For example, it is used separately by “Başkent Doğalgaz Dağıtım A.Ş.” It has at least one spelling as “doğalgaz” on turkiye.gov.tr: “Doğalgaz Dağıtım Şirketlerinin Sunduğu Hizmetler” (Services Provided by Natural Gas Distribution Companies); it also has another spelling written separately on turkiye.gov.tr: “Doğal Gaz Abonelik Başvurusu” (Natural Gas Subscription Application); and it is written as separate words by the Ministry of Energy and Natural Resources. Therefore, the term proves problematic to prove or disprove as a creative spelling, even when considered in the period it was adopted.

3.3. DEFINING NEOLOGISMS

Neologisms found in the data are examined Tweet by Tweet to define their meanings and in what context they are used. This may also prove crucial in identifying and categorising their word formation methods. The following list is the meaning of each neologism found by this study and exemplified with a sentence. However, the example sentences are created for this study and are not taken from the data collected from X, as directly sharing the data is not allowed per the terms and conditions of use of the API.

Afk: is an abbreviation to describe a person who is away from the keyboard and unavailable on the computer. It is an initialism formed from “away from keyboard.”

Example sentence: Ben biraz afk kalacağım, yemek hazırmış.

Akmak: “Akmak” is a word for enthusiastically participating in an activity, such as hanging out or visiting a place

Example sentence: Yarın Bahçeli’ye akalım mı?

Alfa: “Alfa” is a word that describes dominant individuals with leadership abilities.

Example sentence: Oha, adamın yaptığına bak, tam bir alfa.

Anlık: “Anlık” is a word that describes pictures taken at the moment to be sent to someone.

Example sentence: Kanka bana bir anlık atsana.

Aşko: “Aşko” is a new form of “aşkım” (my love) used specifically outside of romantic relationships, usually to refer to friends.

Example sentence: Aşko, yarın sinemaya gideyim diyorum ama sen de gelir misin?

Atar yapmak: Getting extremely angry and showcasing it with one’s behaviours and words.

Example sentence: Ahmet kaza yaptığı zaman çok fena atar yapmıştı.

Banlamak: Permanently barring someone from a platform such as social media, web pages, video games, etc.

Example sentence: Doğa Facebook'ta küfür ettiği için hesabı banlanmış.

Ben şok: A sarcastic way of saying you are shocked without really being shocked.

Example sentence: Bakkalda en sevdiğim çikolata bitmiş, ben şok!

Boomer: "Boomer" usually describes the generation of people, also known as baby boomers, born between 1946 and 1964. However, the word is now used to describe people with outdated worldviews regardless of age or generation.

Example sentence: Boomera bak, telefonların zararlı bir icat olduğunu düşünüyor.

Boş yapmak: Speaking or acting without an aim.

Example sentence: Sen Ahmet'in lafına ne bakıyorsun, boş yapıyor.

Buga girmek: The term is used initially to describe actual computer bugs. However, it is now used to describe situations in which people momentarily freeze and do not know how to act.

Example sentence: Dün gece eve dönerken patronumla karşılaşınca bir anda buga girdim.

Cringe: Being overly embarrassed, mainly due to something another person does.

Example senetnce: Ahmet'in yere çöp attığını görünce acayip cringe oldum.

Çar: Normally used to refer to game characters, çar is now used to describe people as a placeholder for “karakter” or “kişi”.

Example sentence: Ahmet çok garip çar, geçen gün sucuklu tostun arasına cips koyup öyle yiyordu.

Dişil Enerji: A term to describe feminine traits exhibited by a person.

Example sentence: Ayşe bu aralar buram buram dişil enerji yayıyor.

DM: Privately and directly messaging another person, usually on social media.

Example sentence: Abla sana bir DM attım, müsait olunca bakar mısın?

Düşmek: Being attracted to another person.

Example sentence: Buluşmaya gelirken hediye olarak uzun zamandır istediğim boya setini almış, ben direkt bu çocuğa düştüm.

Efso: A new word format for saying “efsane”.

Example sentence: Deadpool v Wolverine filmini izledin mi? Film efso olmuş.

Engel atmak: Blocking all methods of communication with someone on a specific platform such as X, Facebook, or WhatsApp.

Example sentence: Ayşe ayrılır ayrılmaz bana her yerden engel atmış.

E-reçete: Digitalized prescriptions that consist of a set number of characters codified in a certain way.

Example sentence: Doktor e-reçete yazdı ama kodu yazdığı kağıdı ne ben, ne de eczacı okuyabildik.

Eril Enerji: A term to describe masculine energy emanating from someone.

Example sentence: Bu aralar Ahmet'in eril enerjisi bir düşüşte gibi hissediyorum.

Erko: A shortened version of "erkek", used in a derogatory way.

Example sentence: Bu erkolar hep böyle, hiçbir halttan anlamazlar.

Fake hesap: An account, usually on a social media platform, created with fake information to hide the user's true identity or deceive others.

Example sentence: Adamın adı Haydar Trenseveroğulları, fake hesap olduğu çok bariz.

Favlamak: Saving a social media post as a favoured one, allowing one to access it later easily.

Example sentence: Bu maçı kesin 3-2 Türkiye kazanır, favlayıp bekleyin.

Füze atsaydın: A term to describe overkills.

Example sentence: Sevdiğim kız dün Ayşe ile konuşurken benim için çok kısa boylu demiş. Öyle ölmem ya, füze atsaydın.

Ghostlamak: Avoiding someone both in person and on social media platforms.

Example sentence: Ahmet beni geçen haftadan beri ghostluyor, hiçbir mesajıma dönüş yapmadı.

Gizlilik politikası: A legal term to categorise privacy policies, usually of companies, organisations, or governments.

Example sentence: X'in gizlilik politikası güncellenmiş.

Glow up: A person changing considerably in a positive way usually used for appearances.

Example sentence: Çocuğun beş yıl önceki fotoğrafına ve şimdiki haline bakıyorum da acayip glow up yaşamış.

GOAT: An acronym that stands for "Greatest of All Times".

Example sentence: Antep fıstıklı çikolata, bütün çikolataların GOAT'udur.

Gümlemek: Laughing very hard.

Example sentence: Ahmet'in şakasına fena gümledim.

Güno: Shortened version of "günaydın".

Example sentence: Güno, bugün nasılsın?

Halis mi?: A condescending way of questioning the reality of something.

Example sentence: Adamın hırsızlık yapıp üzerine bir de zeytinyağı gibi üste çıkmaya çalışması halis mi peki?

Influencer: A person with a vast following and influence on the internet.

Example sentence: Ahmet influencer olacağım diye tuturmuş, adama seni millet niye takip etsin diye sorunca verecek cevabı yok.

Kanzi: A term to describe alt-right people, usually from generation-z.

Example sentence: Aynen kanzi, hep birlikte el ele verip süper güç olacağız.

Konum atmak: Sending location info through navigation apps such as Google Maps.

Example sentence: Sizin evi bulamadım, konum atsan daha rahat olur.

Köpke: A cute way of saying “köpek

Example sentence: Dışarıdaki köpkeyi gördün mü, çok tatlı.

Manifestlemek: Thoughts and dreams becoming realised in the real world.

Example sentence: Sürekli kaza yaparız diye diye bize kaza yaptırdın, bu kazayı sen manifestledin.

Meme: An amusing item such as a captioned picture or a video.

Example sentence: Dün gece Instragram’da gezerken çok komik bir meme gördüm.

Mezuna kalmak: The term is used to describe students graduated from high school but did not enroll to any university, usually to prepare fo the entrance exams.

Example sentence: Ahmet'in sıralamsı çok kötü değildi ama Hacettepe gelmediği için mezuna kalacakmış, seneye tekrar deneyecek.

Müko: Shortened form of "mükemmel".

Example sentence: Geçen bir ceket aldım, üzerime müko oldu.

Ne münasebo: Shortened way of saying "ne münsaebet".

Example sentence: Ben neden buna tenezzil edeyim ki, ne münsebo.

NPC: NPC is an abbreviation that stands for "Non-Player Characters." This term is usually used for tabletop RPGs or computer games to describe any character in the game that the player does not directly control. However, this study finds that it is used to describe people with dull personalities and no original thoughts.

Example sentence: Ahmet tam bir NPC, adamın hiçbir hobisi yok.

Patlamak: Laughing excessively.

Example sentence: Videodaki köpek kanepeden düşünce patladım.

PC: A shortened version of "Personal Computer" used in Turkish to refer to any computer, usually as a placeholder for the word "bilgisayar."

Example sentence: Benim PC çok eskidi, para biriktirip yenisini almak istiyorum ama laptop mu alsam masaüstü PC mi karar veremedim.

Pick me: A derogatory term describing a person acting in a way to get majority favour from a group by any means necessary.

Example sentence: Ahmet'in Ayşelere yaranmak için kendini ezdirmesi çok pick me bir davranış ve aşırı itici.

PP: Acronym of "profile picture" used in the context of social media profile pictures.

Example: Ahmet yeni pp yüklemiş, gördün mü?

R yapmak: Budging or changing one's perspective or backing away from a situation.

Example sentence: Ahmet garsonla laf dalaşına girmişti ama müdürün geldiğini görünce hemen r yaptı.

Reyiz: A creative spelling of the word "reis", usually used humorously.

Example sentence: Ahmet reyiz bu aralar çok sinirli duruyor.

Roket atsaydın: A term to desribe any type of overkill.

Example sentence: Ayşe'ye hediye gönderdiğim çiçeği kurye ile iade etmiş. Öyle ölmem, roket atsaydın.

Salmak: Giving up on someone or something.

Example sentence: Ahmet evin temizliğini tamamen salmış.

Shiplemek: Act of creating a romantic pair between two individuals that otherwise have no romantic connections.

Example sentence: Şu dizide Ahmet'i ve Ayşe'yi çok fena shipliyorum, bence çok iyi bir çift olurlardı.

Sıfadül eşgal: A term that describes the facial expression of someone when they meet with a condition.

Example sentence: Sabah uyanıp kahvemi bile içmeden işe gidince benim sufadül eşgal de aynı bu çocuğunkine benziyor.

Sigma: Used to describe men with free spirits and

Slay: An expression to praise the appearance or acts of another person.

Example sentence: Yeni elbisen çok güzel olmuş, slay.

Stalklamak: Act of repeated surveillance from a person or group towards another one on social media.

Example sentence: Ahmet dün gece sosyal medyada eski sevgilisini stalklamış.

Story: Short videos or pictures shared on social media for a set amount of time.
Example sentence: Ahmet, İzmir gezisini storysine atmış.

Sunroof kız: A woman who ties her headscarf in a way that leaves the upper-front part of her hair exposed.

Example sentence: Ayşe de sunroof kız, örtüsünü bağlarken saçının bir kısmını hep bilerek açık bırakıyor.

Takipçi kasmak: Act of increasing one’s social media following through various methods.

Example sentence: Ahmet takipçi kasmak için köpeğinin videolarını çekip internete yüklemeye başlamış.

Tilt olmak: Suffering an immense frustration.

Example sentence: Şu oyunda 5. seviyeyi bir türlü geçeemediğim için acayip tilt oldum.

Triggerlanmak: Being extremely offended by a situation.

Example sentence: Ahmet’in insan hakları hakkındaki düşüncelerine triggerlandım.

Vibe: The feelings someone, something, or somewhere makes others feel without making a conscious effort.

Example sentence: Ahmet’in yeni arkadaşından çok iyi bir vibe aldım.

Yargı dağıtmak: The expression used to identify situations in which someone puts another person in their place, gives them a piece of one's mind, or puts them down.

Example sentence: Ahmet’in maaşı yatmayınca Ahmet iş yerinde yargı dağıtmış.

Yeto: A shortened version of “yeter”.

Example sentence: Yeto, canım çok sıkıldı, biraz farklı bir şeyler yapalım.

Yıkık: A derogatory term to refer to people in a self-induced embarrassing state.

Example sentence: Ahmet, arkadaşına rest çekip küstükten sonra gidip yalvararak barışmaya çalıştı, yıkık bir hareket.

Youtuber: An individual who prepares videos to be published on his YouTube channel.

Example sentence: Tier Zoo, en sevdiğim Youtuber olabilir. Hem eğlenceli, hem de öğretici içerikleri var.

Yükselmek: Being overly attracted to someone, especially on a physical level.

Example sentence: Ahmet'in yeni stilini görünce biraz yükseldim.

Yürümek: Fliriting with someone.

Example sentence: Ahmet'e yürüyorum ama bana karşılık vermiyor gibi.

Zırvana: Combined "zırvalamak" and "nirvana", meaning blather's epitome or top point.

Example sentence: Ahmet bu aralar zırvanada, kendisini dünyadaki en önemli kişi sanıyor.

3.4. CATEGORIZING NEOLOGISMS

According to Fang (2021), neologisms can be categorised in four different ways: function, coinage process, word formation process, and source. This section of the study will analyse the neologisms in all categories and make further suggestions on previously suggested categories by Yurtbaşı (2017), as some of

the word formation processes observed in the analysis are not adequately explained by Yurtbaşı's word formation processes for neologisms.

3.4.1. New Formation Processes for Neologisms

The study observes five different formation methods for neologisms that previous studies have not adequately identified. It should be noted that while neologisms have overlapping formation methods with word formations, not all neologism formation methods are considered as word formation methods since neologisms can be any number of strings of words. These methods are blending, hybrid neologisms, initialisms, acronyms, phraseological neologisms, and hypocoristic neologisms. This section of the study will explain the new word-formation processes for neologisms to clarify the categorisation of the words based on the word-formation processes.

3.4.1.1. Blending

Blending in morphology is defined as partially removing at least two words. The difference between compounding and blending comes from clippings. Blendings are created by removing some of the morphemes of at least one of the words used in the process. Morphologically, blendings can be examined in two ways: partial blending and full blending.

Partial blending is made by only clipping one word while leaving another completely intact. An example of this type of blending would be "e-reçete". The word is created by blending the words "elektronik" and "reçete" by clipping "elektronik" and leaving "reçete" intact. Full blendings are created by clipping both words into one new form. "Zırvana" is an example found in this study for neologisms created in such a manner as it clips both "zırva" and "nirvana" into one new form.

Blendings can also be overlapping or non-overlapping. Overlapping blends are created by combining words that have partial overlapping consonants, wovels, or syllables. “Zırvana” is an example of such blendings found in this study as “zırva” and “nirvana” have overlapping syllables, -va, at the end of one word and in the middle of another. Non-overlapping blendings are the contrary of overlapping blendings, i.e. the combined words share no overlaps. “E-reçete” is an example for non-overlapping blendings as “elektronik” and “reçete” are not blended in a way with overlaps.

Finally, blendings can be attributive or coordiante. Attributive blends have one part, which is the head, and the attributive part. There are no examples of attributive blending in this study; however, an example from English can be “porta-light”. Porta-light is used to define portable light; it does not refer to light portability. Therefore, *light* is the head, and *portable* is the attribution here. On the other hand, coordinate blendings carry all parts of the blends in an equal manner and have two heads. “Zırvana” is an example of coordinate blendings as “zırvana” is neither just “zırva” nor “nirvana” but a total and equal combination of both, reaching the nirvana of blather.

3.4.1.2. Initialisms and Acronyms

Initialisms and acronyms are created through the same process but differ in pronunciations. Both word formations are created by using the parts of the phrases they are created out of. The difference is that acronyms are pronounced as separate words, whereas words created through initialism are pronounced as individual letters. A typical example of an initialism in Turkish would be “TBMM” (Türkiye Büyük Millet Meclisi), pronounced as individual letters. On the other hand, NATO (North Atlantic Treaty Organization) would be an example of an acronym as it is pronounced as a separate word. These types of neologisms are not observed to be directly formed from Turkish words in the data but are rather borrowings from foreign languages.

3.4.1.3. Hybrid Neologisms

Hybrid neologisms are a new word formation process for neologisms that have not been proposed for Turkish before. Hybrid neologisms are created by blending two languages to create one new word or a string of words. This study finds this is common in Turkish as most borrowings either get affixes or a supplementary Turkish word next to it. One of the examples of these neologisms would be “stalklamak”. “Stalklamak” is created through the borrowed word “stalk” from English and has two suffixes -la and -mak.

3.4.1.4. Phraseological Neologisms

Phraseological neologisms are newly created strings of words that create a conventional and fixed use of phrases longer than two words. One of the examples from this study of phraseological neologisms would be “roket atsaydın” and “füze atsaydın”. Since Turkish is a pro-drop language, the subject is omitted from the phrase. The verb conjugation clearly indicates the second-person singular pronoun as the subject of the phrase, which is evident from the suffix -dın. Since both phrases drop the subject but are implicit from the verbs, these phrases are constructed with three words rather than two and are considered examples of phraseological neologisms.

3.4.1.5. Hypocoristic Neologisms

Hypocoristic forms refer to words morphologically transformed into forms denoting affection. Such as adding -y / -ie into English words such as “kitty” or “plushie”. This study proposes that neologisms can be created through hypocoristic means.

There are various neologisms found in this study that are made by clipping a word and adding -o at the end. “müko”, “güno”, “aşko” “ne münsaebo”, “efso” and

“erko”. All the neologisms, except for “erko”, follow a pattern of endearment or affection to the meaning and are created with specific rules.

Firstly, all these words are clipped from the end. The most notable pattern here is clipping every letter except for the first three letters, except for “ne münsabo”. In addition, the letter -o is added to the end of all clipped words to create a shortened and more affectionate way of saying the same thing. In a way, these neologisms do not add anything but a sentiment of affection to the words created through this process. The only exception for this is “erko”, though it is unclear if the word “erko” is created before or after the other neologisms. If it is created before or after other neologisms, it may indicate a shift in the semantics associated with the process. Alternatively, suppose it is created with other neologisms, i.e. before and after some of them. In that case, it may only be an exception to the rule since this seems like a relatively new concept of forming neologisms.

This neologism formation process may be attributed to a relatively rare hypocoristic form observed in Turkish. Adding the suffix—o to proper nouns may add an affection to the meaning, as can be observed in examples such as “Hamido” and “Yiğido.” However, the new process does not seem to be restricted to proper nouns.

3.4.2. Neologisms and Their Categories

This section will categorise the neologisms found in this study based on their function, coinage process, formation process, and source. The categories are based on Fang’s (2021) definitions, and the word formation processes are identified using Yurtbaşı’s (2017) work with the addition of new neologism formation processes defined in *3.4.1*.

1. Afk

- Function: It is a neologism with expressive function as it does not fill a scientific gap but adds new forms to discourse.
- Coinage process: It is a borrowing for the Turkish language since the word originates from English.
- Formation process: The word is formed through initialisms in English, as all letters are pronounced separately. It should be noted that this neologism is written with the Turkish pronunciation of the letters rather than the English pronunciation. This is a borrowed word for Turkish.
- Source: It is also an imported word for Turkish regarding the source. However, it is a pop culture source for English.

2. Akmak

- Function: It is a neologism with expressive function as it does not fill a scientific gap but adds new forms to discourse.
- Coinage process: Its formation is an example of semantic shift, changing the original meaning of the verb “akmak”.
- Formation process: The formation process is a semantic shift as it changes the semantics of the verb “akmak” in Turkish.
- Source: The source can be attributed to popular culture as it is a highly informal word with no ties to other sources.

3. Alfa

- Function: Originally, alfa was a referential word that was used to describe various personalities of people.
- Coinage process: The word is coined through borrowing for Turkish.

- Formation process: The word is borrowed into Turkish, but it is initially formed as a hybrid neologism, borrowing the first letter of the Greek alphabet and turning it into an adjective.
- Source: Regarding its source, alfa is proposed as a scientific word even though the notion seems to be losing a fraction.

4. Anlık

- Function: It is an expressive neologism as it adds new forms to the discourse.
- Coinage process: It is a semantically shifted word.
- Formation process: Anlık is formed by semantically shifting the original word.
- Source: Popular culture is the source of the word, which is an informal use originating from social media.

5. Aşko

- Function: An expressive word adds a new form to the discourse.
- Coinage process: It is a new word created through the hypothesised hypocoristic neologism formation method.
- Formation process: The word is formed through the hypothesised hypocoristic neologism formation process by clipping the word except for the first three letters and adding the letter -o to the end to create a sense of endearment.
- Source: The word’s source can be traced to popular culture

6. Atar yapmak

- Function: An expressive word adds a new form to the discourse.
- Coinage process: It is a newly created word.

- Formation process: It is created through combining the noun “atar” with the verb “yapmak”.
- Source: It is a word born out of popular culture.

7. Banlamak

- Function: It is a referential word as it fills a specific gap in a specific scientific field, i.e. computer sciences. Initially, the word was expressive and not limited to computer science, but it was used explicitly for software-related bannings in Turkish.
- Coinage process: It is a borrowed word from English.
- Formation Process: A hybrid neologism borrowed from English and affixed with Turkish suffixes.
- Source: Scientific word for Turkish, popular word for English.

8. Ben şok

- Function: It is expressive as it adds new forms to the discourse.
- Coinage process: It is a newly created expression that combines two words.
- Formation process: It is formed by combining two words to denote a new meaning.
- Source: The source of the word can be traced back to popular culture.

9. Boomer

- Function: It is an expressive word for Turkish as it adds a new form to the discourse. However, the original word is a reference word used in social studies to identify a certain generation.
- Coinage process: It is a borrowed word from English.
- Formation process: “Boomer” is a borrowed word with no additions.

- Source: The word was originally a scientific word, but its use in Turkish is popular culture since it has an entirely new meaning compared to the scientific meaning.

10. Boş yapmak

- Function: It is an expressive term as it adds a new form to the discourse.
- Coinage process: It is created as a new word.
- Formation process: The word is created by combining two words together.
- Sources: It is a word born out of popular culture.

11. Buga girmek

- Function: An expressive term adds a new form to the discourse.
- Coinage process: It may indicate a newly created word since it is a hybrid. However, it is coined through semantic shifts since the expression is unrelated to actual computer bugs.
- Formation process: It is a hybrid neologism created by borrowing “bug” from English, adding a Turkish suffix, and combining it with a Turkish verb.
- Source: It is a word born out of popular culture.

12. Cringe

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a borrowed word from a foreign language.
- Formation process: It is a direct loanword, or borrowing, originating from English.
- Sources: It is a word born out of popular culture.

13. Çar

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word may seem like a borrowed word as it is created through a long process. Initially, it is a loanword born from the clipped version of the English word “character”. This is a direct transliteration process as the word is also clipped as “char” in English. However, this borrowing was used in Turkish, especially in the context of computer games. The definition found in this study suggests a semantic shift in the previously coined word.
- Formation process: It is a borrowing for Turkish and clipping for English. However, the original borrowed meaning experiences a semantic shift, thus making this a neologism formed through the semantic shift.
- Sources: The word’s source can be traced to popular culture.

14. Dişil enerji

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is borrowed from English.
- Formation process: More specifically, the word is the calque of the English word “feminine energy”.
- Source: The word’s source can be traced back to an imported word.

15. DM

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is coined through borrowing.
- Formation process: It is a borrowed word from English originally formed through initialism.
- Source: The word’s source in Turkish can be traced back to an imported word, a popular culture-sourced word.

16. Düşmek

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: Semantic shift.
- Formation process: The expression is created through a semantic shift from the verb “düşmek”, which originally means “to fall”. The formation process may be related to the expression “falling for someone” in English; however, there were no indications to prove or disprove this notion.
- Sources: The word’s source can be traced to popular culture.

17. Efso

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a newly created word.
- Formation process: The word is formed through the hypothesised hypocoristic neologism formation process by clipping the word except for the first three letters and adding the letter -o to the end to create a sense of endearment.
- Sources: The word’s source can be traced to popular culture.

18. Engel atmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The expression is created as a newly coined expression.
- Formation process: The expression is created by combining two words without majorly altering their meanings but by focusing them on a specific context.
- Sources: The word's origin can be traced back to popular culture.

19. E-reçete

- Function: It is a referential word as it fills a gap in medical sciences.
- Coinage process: It is a newly created word.
- Formation process: The word is formed by blending. It is a partial blending as “reçete” is not clipped, it has no overlaps between the words “elektronik” and “reçete”, and it is also a coordinate blending since both parts are heads of the word.
- Sources: It is a scientific word originating from medical sciences.

20. Eril enerji

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a borrowed expression.
- Formation process: It is formed through borrowing via calque of the English word “masculine energy”.
- Sources: The word's origin can be traced to a foreign word; therefore, it is imported. However, the original word is a popular culture word.

21. Erko

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a newly created word
- Formation process: The word is created through a hypothesised hypocoristic neologism formation process by clipping the word except for the first three letters and adding the letter -o at the end to create a sense of endearment.
- Sources: The word's origin can be traced back to popular culture.

22. Fake hesap

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a newly created word since “fake” has had a derogatory connotation in Turkish for quite a while.
- Formation process: The word is a hybrid neologism since the word “fake” is borrowed from English and combined with the Turkish word “hesap.”
- Sources: The word's origin can be traced back to popular culture.

23. Favlamak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The term is a new word created.
- Formation process: The word is created by clipping the English word “favourite”, reducing it to the first three letters and adding Turkish suffixes. Therefore, it is a hybrid neologism.
- Sources: The term’s origin can be traced to popular culture.

24. Füze atsaydın

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is coined by creating a new neologism.
- Formation process: The word is formed as a phraseology since it exhibits pro-drop features explained in 3.4.1.4.
- Sources: The origin of the word can be traced to popular culture.

25. Ghostlamak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is coined through borrowing.

- Formation process: The word is formed by adding Turkish suffixes to the word “Ghost,” which is borrowed from English. Therefore, it is a hybrid neologism. It seems to imitate the English neologism “ghosting directly” but changes the word's inflexions with Turkish alternatives.
- Sources: The word's origin can be traced to an imported word from English.

26. Gizlilik politikası

- Function: It is a referential word as it fills a gap in the field of law.
- Coinage process: The expression is coined through borrowing.
- Formation process: The expression is a borrowed word, a calque, directly taken from the English expression “privacy policy”.
- Sources: The source of this neologism can be attributed to science.

27. Glow up

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The expression is coined by borrowing.
- Formation process: The expression is a borrowing in Turkish and an example of a semantic shift in English.
- Sources: The word is an imported word for Turkish. It originates from popular culture in English.

28. GOAT

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a borrowed word in Turkish.
- Formation process: The word is a direct borrowing; however, the original English word is an acronym created out of “greatest of all time”.
- Sources: The source of the word is an imported word for Turkish, and the English word can be traced to popular culture.

29. Gümlemek

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The expression is a semantic shift.
- Formation process: The expression is formed through semantic shifting.
- Sources: The origin of the word can be traced to popular culture.

30. Güno

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The expression is a newly created word.
- Formation process: The formation process is the hypothesised hypocoristic neologism. The original word “günaydın” is clipped except for the first three letters, and an -o is added to the end to create a sense of endearment.
- Sources: The origin of the word can be traced to popular culture.

31. Halis mi?

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is an expression created through semantic shift.
- Formation process: The expression itself is a semantic shift. The original form and meaning are preserved, but now, it also expresses a condescending view.
- Sources: The origin of the word can be traced to popular culture.

32. Influencer

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is borrowed from English and is a newly created expression in English.

- Formation process: The word is used as a loanword in Turkish. The English word is created with derivation.
- Sources: The word's source can be traced to an imported word for Turkish, and it is a popular culture term in English.

33. Kanzi

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a new expression.
- Formation process: The word may have two origins. The most likely option is a phono-semantic shift from “kanka”. However, it is also likely to be attained by a famous intelligent monkey named Kanzi. The research has no definitive conclusions on the origin. However, it suggests a greater likelihood of a phono-semantic shift as the word “kanka” was shifted multiple times before “Kanka” itself is a blending of “kan” and “kardeşi”i. The new word then experiences multiple phono-semantic shifts throughout the time with examples such as “kanki” and “panpa”; however, these two phono-semantic shifts seem to be not used widely as per the analysis of this study as “panpa” never occurs in the data and “kanki” only occurs once. Therefore, it is likely that “kanzi” is just a phono-semantic shift from “kanki”, just as in the previous examples.
- Sources: The origin of the word can be traced to popular culture.

34. Konum atmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: This is a newly coined expression.
- Formation process: The expression combines “konum” and “atmak”. It does not change the relative meanings of the words but uses a rather unique verb in a context where “göndermek” would be more appropriate.
- Sources: The origin of the word can be traced to popular culture.

35. Köpke

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a new expression.
- Formation process: The formation process indicates a phono-semantic shift. It is a creative way of spelling or articulating the word “köpek” in a way that sounds more endearing, possibly to avoid negative connotations associated with the word “köpek”.
- Sources: The origin of the word can be traced to popular culture.

36. Manifestlemek

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a borrowed form in Turkish.
- Formation process: The term itself is a hybrid neologism, a directly borrowed form from English, used with Turkish suffixes.
- Sources: The origin of the word can be traced to popular culture.

37. Meme

- Function: It is an expressive word that adds new functions to the discourse. However, Charles Darwin coined the word as a scientific word, making it a referential term in the past for English.
- Coinage process: The word is coined through borrowing.
- Formation process: The word in Turkish is a direct borrowing. However, it is a self-coined term that experienced a semantic shift in English.
- Sources: The word's origin, for Turkish, can be traced to popular culture as the meaning used in Turkey conveys “internet memes”, i.e. short and funny pictures and videos. The word itself is a scientific word in English.

38. Mezuna kalmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a newly created expression.
- Formation process: The word is created by combining “mezun” and “kalmak”.
- Sources: The origin of the word can be traced to popular culture.

39. Müko

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly coined term.
- Formation process: The word's formation process aligns with the hypothesised hypocoristic neologism formation process. It clips the letters of “mükemmel” except for the first three letters and adds the letter—o to the end as a form of endearment.
- Sources: The origin of the word can be traced to popular culture.

40. Ne münsebo

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly coined term.
- Formation process: The word formation process indicates similarities with the hypothesised hypocoristic neologism formation process. It does not clip every letter except the first three letters, but it still adds—o at the end as a sense of endearment. A change in the clipping rule may result from it being a combined word rather than one word.
- Sources: The origin of the word can be traced to popular culture.

41. NPC

- Function: It is an expressive word that adds new functions to Turkish discourse. However, the original term in English is a referential word that covers certain coded characters in software.
- Coinage process: It is a direct borrowing in Turkish and a newly created form in English.
- Formation process: It is a direct borrowing in Turkish but is constructed through initialism in English.
- Sources: The word's origin can be traced to popular culture in Turkey and software engineering in English.

42. Patlamak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: It is a semantically shifted word.
- Formation process: The word is semantically shifted to express excessive laughter.
- Sources: The origin of the word can be traced to popular culture.

43. PC

- Function: It is an expressive word that adds new functions to the discourse. However, the original word in English is a referential word covering a newly invented concept, i.e. personal computers.
- Coinage process: The word is a direct borrowing in Turkish.
- Formation process: The word is directly borrowed in Turkish and constructed through initialism in English.
- Sources: The word's origin can be traced to popular culture in Turkey, but it is a scientific word in English.

44. Pick me

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through combination.
- Sources: The origin of the word can be traced to popular culture.

45. PP

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through combination.
- Sources: The origin of the word can be traced to popular culture.

46. R yapmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly created term.
- Formation process: The word formation is somewhat unique. “R”, in this expression, signals the “R” gear in cars, i.e. reverse gear. However, “R” is not used in this sense in English; rather, it is a direct initialism borrowed from an English word. Considering that initialism is combined with Turkish words, the word can be said to have been formed through a hybrid neologism formation process.
- Sources: The origin of the word can be traced to popular culture.

47. Reyiz

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly created expression.
- Formation process: The word is created through a phono-semantic shift by adding the letter -y to the word “reis”.
- Sources: The origin of the word can be traced to popular culture.

48. Roket atsaydın

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly created expression
- Formation process: The word is formed as a phraseology since it exhibits pro-drop features explained in 3.4.1.4.
- Sources: The origin of the word can be traced to popular culture.

49. Salmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is coined through a semantic shift.
- Formation process: The word is semantically shifted from its prior definition.
- Sources: The origin of the word can be traced to popular culture.

50. Shiplemek

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English and used with Turkish suffixes. Thus, it is formed through a hybrid neologism process.
- Sources: The origin of the word can be traced to popular culture.

51. Sıfadül eşgal

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word goes through a semantic shift to gain a new definition.
- Formation process: The word is a semantically shifted archaic expression.
- Sources: The origin of the word can be traced to popular culture.

52. Sigma

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is borrowed from Greek.
- Sources: The word's origin can be traced to popular culture in Turkish. However, the word in English is rooted in social sciences.

53. Slay

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English.
- Sources: The origin of the word can be traced to popular culture.

54. Stalklamak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.

- Formation process: The word is directly borrowed from English, and the English word is formed through a semantic shift. The word also includes Turkish suffixes. Therefore, it is a hybrid neologism.
- Sources: The origin of the word can be traced to popular culture.

55. Story

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through semantic shifting.
- Sources: The origin of the word can be traced to popular culture.

56. Sunroof kız

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly coined term.
- Formation process: The word “sunroof” is directly borrowed from English and combined with the Turkish noun “kız” to create a hybrid neologism.
- Sources: The origin of the word can be traced to popular culture.

57. Takipçi kasmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly coined expression.
- Formation process: The word is formed by combining “takipçi” and “kasmak”.
- Sources: The origin of the word can be traced to popular culture.

58. Tilt olmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through semantic shifting. The expression also combines the English worth with the Turkish verb “olmak”, thus creating a hybrid neologism.
- Sources: The origin of the word can be traced to popular culture.

59. Triggerlanmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through semantic shifting. The expression also combines the English worth with the Turkish suffixes, thus creating a hybrid neologism.
- Sources: The origin of the word can be traced to popular culture.

60. Vibe

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through semantic shifting.
- Sources: The origin of the word can be traced to popular culture.

61. Yargı dağıtmak

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a newly coined expression.
- Formation process: The word is formed via combining “yargı” and “dağıtmak”.
- Sources: The origin of the word can be traced to popular culture.

62. Yeto

- Function: It is an expressive word as it adds new functions to the discourse.
- Coinage process: The expression is a newly created word.
- Formation process: The formation process is the hypothesised hypocoristic neologism. The original word “yeter” is clipped except for the first three letters, and an,—o is added to the end to create a sense of endearment.
- Sources: The origin of the word can be traced to popular culture.

63. Yıkık

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is coined through a semantic shift.
- Formation process: The word is semantically shifted from its prior definition.
- Sources: The origin of the word can be traced to popular culture.

64. Youtuber

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is a direct borrowing.
- Formation process: The word is directly borrowed from English, and the English word is formed through semantic shifting.
- Sources: The origin of the word can be traced to popular culture.

65. Yükselmek

- Function: It is an expressive word as it adds new functions to the discourse.
- Coinage process: The word is coined through a semantic shift.
- Formation process: The word is semantically shifted from its prior definition.
- Sources: The origin of the word can be traced to popular culture.

66. Yürümek

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is coined through a semantic shift.
- Formation process: The word is semantically shifted from its prior definition.
- Sources: The origin of the word can be traced to popular culture.

67. Zırvana

- Function: It is an expressive word that adds new functions to the discourse.
- Coinage process: The word is newly coined neologism.
- Formation process: The word is created through blending. The words “zırva” and “nirvana” are combined together to create this expression. This blending is a full blending as both words are clipped; it is an overlapping blending since they share a syllable, and it is a coordinate blending since both words are heads in the new form.
- Sources: The origin of the word can be traced to popular culture.

3.4.2.1. Frequency of Neologisms per Category

This section will analyse the frequency of neologisms in each category and identify the most frequent processes to create neologisms in Turkish per the data available for this study. The following table shows the percentage of neologisms based on function, coinage process, formation process, and source. The data is

analysed according to how the word is formed and used as per the Turkish language and does not account for previous forms or foreign versions of the neologisms. Thus, a borrowed word from English with the formation process of initialism, for example, will be categorised as a borrowed word but not as an initialism.

Table 3. *Neologism Frequencies Based on Function*

Functions	Frequency
Expressive	**95.52**
Referential	**4.48**

The table shows a massive difference between the two functions to categorise neologisms. The data suggests that the vast majority of Turkish neologisms on social media are expressive words formed to introduce new forms to the discourse and are not formed to fill gaps in specific gaps in specialised fields.

Table 4. *Neologism Frequencies Based on Coinage*

Coinage	Frequency
Semantic shift	16.42
Borrowing	40.30
New words	43.28

The data suggests that the majority of Turkish neologisms on social media are formed as new words, with 43.28%. This may suggest that neologisms on social media are created or used to express concepts in a shorter way, which can be evident from examples. Borrowings with 40.30% follow this, almost a difference of 3%. All the borrowings found in the data are taken from English, even if the English word is a borrowing itself. This may suggest that the fast spread of information on social media facilitates borrowings from other languages. English is the primary source for such borrowings as it is the most spoken second language in Turkey and the majority language of the internet. Finally, semantic shifts are the least used type of coinage with 16.42%, a sharp decrease from

borrowings and new words. Nevertheless, the number of semantic shifts found in the data is not negligible.

Table 5. *Neologism Frequencies Based on Formation*

Neologism	Researchers
Blending	2.99
Borrowing	26.87
Calque	4.48
Clipping	0.00
Combining	11.94
Compounding	0.00
Hybrid	17.91
Hypochorostic	10.45
Phono-semantic	4.48
Phraseological	2.99
Semantic shift	17.91

The data shows that borrowings are the most common way of forming new neologisms on social media, 26.87% of all neologisms found in this study. This may result from globalisation and the fast spread of information through the Internet. Novel concepts and ideas are spread fast on the internet, and with the increasing population of bilinguals all across the world, if an English word exists to fill either an expressional or referential need in discourse, speakers may opt to borrow it directly from English as other bilingual native speakers would understand it. All of the borrowings in this study taken from English support this idea.

The second most common ways of creating neologisms are hybrid neologisms and semantic shifting, both accounting for 17.91% of the data. Hybrid neologisms may include other aspects in themselves, such as derivation (Ghostlamak) or initialism (R atmak). However, what defines hybrid neologisms is that they are formed by combining two languages. English is the source language for all hybrid neologisms found in this study.

Semantic shifts are created through repurposing old lexical units with new definitions. The data suggests that this mainly occurs in Turkish neologisms on social media as a humour enhancer. Words such as “patladım”, “gümledim”, “halis mi”, and “sıfadül eşgal” are semantically shifted to achieve a comedic effect. However, this is not the sole purpose, as it is also evident from words such as “akmak” that semantic shifts occur for other reasons, too, even though all of them are highly informal in nature.

The fourth most common way of forming neologisms on social media for the Turkish language is by combining words. This method, suggested by Yurtbaşı (2017), requires two words to be used as one lexical string. TDK suggests that this is a way of compounding words, as they divide compounded words into two categories: bitişik yazılan bileşik kelimeler (solid compounds) and “ayrı yazılan bileşik kelimeler” (spaced compounds). However, Yurtbaşı suggests to make a distinction between the two methods. This may be why no compound words (i.e., solid compound words) were found in the data. It may further suggest that spaced compounds, or combined words, are the more dominant way of compounding on social media for the Turkish language.

The following most common way to form neologisms is the hypocoristic neologisms suggested in this study. There seem to be three patterns for forming hypocoristic neologisms. First, the words are clipped. This is observed in all examples without an exception. It should be noted that all examples are clipped as its first three letters, except for “ne münasebo”, so the rule may be clipping rather than clipping the first three letters. The second rule, which can be observed in all examples without an exception, is the addition of the letter -o to the end of the clipped word. Moreover, all examples except for “erko” induce a sense of endearment to the word's meaning, but “erko” is used as a derragotary term. The reason behind this difference is unclear as it was impossible to identify the exact origin of the neologism or establish a timeline for creating hypocoristic neologisms in Turkish.

Phono-semantic formations are following on the list. Initially, this neologism formation is not among the suggestions of Yurtbaşı (2017). However, it is a word formation method coined by Kara (2011). This method means a shift in semantics created by a phonological change. The data suggests that this word-formation process is also used for neologisms. It is observed on a total of three neologisms, and there is not a visible pattern in the phonologic or semantic changes (“Kanka” or “kanki” to “kanzi”; “köpek” to “köpke”; and “reis” to “reyiz”). Two of the words are derragotary terms, while one of them is endearing. Derragatory neologisms have a phonological change from the letter -s to -z or -k to -z. However, there is no clear pattern to establish a link between the two.

Calque neologisms, or translated neologisms as Yurtbaşı puts it, are as prevalent as phono-semantic neologisms. It should be noted that all calque neologisms were borrowed from English.

The least common formation methods were blending and phraseological formations. It should be noted that these neologisms are not only less prevalent in terms of formation methods, but they also score a low-frequency point across all neologisms found in this study. One of the interesting details to note here is on “e-reçete”. Clipping “elektronik” as “e” and adding a hyphen is a previously observed method for creating neologisms in Turkish, as evident from words such as “e-devlet”, “e-okul”, “e-imza” etc., These neologisms were proposed by Tahioroğlu et al. (2017), and they were added to TDK’s dictionary. However, “e-reçete” is also one of the neologisms found by Tahiroğlu et al., but it is not in the dictionaries yet. This is a specific way of creating blended words, which is still used in 2023.

The following method on the list is phraseological neologisms. Again, these are all created for a humorous effect. The study found that all of the phraseological neologisms in Turkish were formed with pronoun drops. All phraseological neologisms may essentially look like two separate lexical units; however, the

subject of the phrase is removed from the phrases since Turkish is a pro-drop language.

Compound neologisms are not found in the data. However, compound words were still common since spaced compound words are categorised under a different name as combined *neologisms.*

The study did not find any neologisms created through clipping. Clipping is a method used to create neologisms such as “lab” (laboratuvar) and büt (bütünleme sınavı). However, even though no neologisms were created through clipping, it was still used as a partial method for creating other neologisms. Hypochorostic neologisms all require clippings; they are not considered clipping because clippings do not have any additions, while hypochorostic neologisms all have an additional letter: -o. In addition, a semantic shift is never observed in neologisms created through clipping. However, hypocoristic neologisms all have a semantic shift that is added by adding one letter at the end of the clipped words. This situation is similar to blending, as blending also requires partial or full clipping, but they are categorised differently since they also experience a semantic change.

Table 6. *Neologism Frequencies Based on Sources*

Neologism	Source
Science	5.97
Popular culture	94.03
Nonce	0
Political	0

An interesting but not surprising finding is that neologisms originating from popular culture are vastly dominant over other sources, with 94.03% of all neologisms found in this study being sourced from popular culture. It is not a surprising finding, considering social media is a highly informal communication medium and not primarily used for political or scientific discussions. Nonce words

were also not expected to be found in the study since nonce words occur once for specific occasions such as literary acts, and any word that occurred less than five times in the whole data set was excluded from the analysis to eliminate spelling errors or similar mistakes. Neologisms with scientific origins amount to a total number of four; however, with the exception of “e-reçete” neologisms with scientific origins may experience a semantic shift and move to the popular culture category. This is an observed trend in neologisms such as “meme” (originally a scientific word proposed by Charles Darwin).

CONCLUSION

This study consists of analysing an eventual corpus compiled from the social media website X. The corpus amounts to 327.262 Tweets with a total word count of 2.463.075. The corpus was limited to Tweets sent from Turkey in the Turkish language between 01.01.2023 and 31.12.2023. The corpus was analysed to find neologisms on Turkish social media, and 74 neologisms were found through analysis. Social media is chosen as a medium to gather spontaneous interactions and to eliminate the observer's paradox.

The study required a Turkish social media corpus that was not available at the time of its writing. Existing corpuses had various problems. Namely, they were too old to use in neologism detection, had a smaller quantity of data than desired, and were not randomly selected. Therefore, a completely new corpus needed to be compiled for the study. The corpus was collected through X API using Python via *Tweepy,* a Python library to access X API. The Tweets are collected through random sampling with four criteria: date, language, location, and geographical origin. The data was cleaned and stored using Python via *Pandas,* a Python library that manipulates and stores data. Repeating Tweets, Retweets, Tweets without the proper geolocation (i.e. Tweets sent from unknown origins or outside of Turkey), and Tweets in languages other than Turkish.

The initial question of the study was to identify the most frequent neologisms used on Turkish social media based on Grive et al.'s framework. This required a meticulous analysis of the corpus. This required an analysis of word frequencies per a set number of words in the data. All the textual data from Tweets had to be tokenised to achieve this. The tokenisation process was conducted through *TRNLP*, a Python library designed for tokenisation and morphological analysis for the Turkish language. However, this library was only used for the tokenisation process. All text data was tokenised to units consisting of one word. This data had to be morphologically analysed for lemmatisation since the analysis accounted for inflexions, i.e. inflected words were grouped together rather than

being accounted as separate units. This process was conducted through *TrMorph, a* library to analyse the Turkish language morphologically. The resulting data was manually grouped according to lemminizations and semantic shifts.

The resulting data was manually analysed to see if they could be considered as neologisms. This required them to be relatively new lexical units with no dictionary entries per the neologism definition. This was a manual process enhanced by *TrMorph, considering the library automatically gives definitions of the words by* looking them up on TDK's online dictionary. This feature helped to speed up this process. However, it was not enough to be wholly relied upon due to the nature of the data. *TrMorph* is a library that can scan the dictionary, but it is not an AI, which would have its own problems, that can analyse words based on context. It only scans singular entries through the dictionary. However, the analysis needed to account for semantic shifts, and this proved manually analysing the data to detect such occurrences. Possible neologisms detected after this process were then analysed according to Grieve et al.'s framework, and neologisms with relevant frequencies were listed.

The findings of the analysis suggested that a total of 67 neologisms were used on Turkish social media based on the available corpus. These neologisms were then listed according to their frequencies on the data set. The most frequent neologisms among the findings of this study were "kanzi" (7.24 frequency), "DM" (6.94 frequency), and "meme" (6.33 frequency). The least frequent neologisms were "boomer" (1.02 frequency), "gizlilik bolitikası" (1.02 frequency), and "ne münasebo" (1.02 frequency).

The findings of the initial question were then analysed to see if any findings matched the neologisms proposed in previous studies by Tahiroğlu et al. (2017), Çokol (2020), and Safa & Bilginsoy (2020). This was to establish whether these neologisms were still in use on social media and may be a possible candidate for being a dictionary entry, considering they would be in use for a considerable time. A total of 21 neologisms were identified to be matched with the findings of

previous studies: 4 matches with Tahiroğlu et al., nine matches with Çokol, and nine matches with Safa & Bilginsoy).

The data was then analysed manually to identify the definitions of all neologisms found in this study. This was a completely manual process conducted by analysing every textual data containing at least one of the neologisms found in the study. This process was handled as how lexicographers write dictionary entries for dictionaries such as *Merriam-Webster.* The process requires the analysis of a word in the relevant context to determine its definition. The definitions for each word were listed in one example sentence. However, due to X's terms and conditions, the example sentences are not given from the corpus but written down by the researcher specifically for this study.

Finally, all neologisms were analysed to be categorised based on their function, coinage, formation, and source with the framework set out by Fang (2021). This process required further analysis of all neologisms to determine relevant information for their categories. Neologism formation was categorised according to Yurtbaşı's (2017) categories of neologism formations in Turkish. However, it was supported by one additional word formation process proposed by Kara (2011), the phono-semantic shift, that Yurtbaşı did not list. Moreover, this analysis concluded that four more neologism processes for the Turkish language are not listed before, 3 of which can be observed in numerous other languages such as English or French. These processes are identified as blending (partially or fully clipping two separate words and compounding them), hybrid neologisms (neologisms created through combining elements of two languages), and phraseological neologisms (phrases consisting of lexical strings longer than three words). In addition, the analysis suggests a process unique to Turkish for creating neologisms. This process is called hypocoristic neologism formation. Typically, hypocoristic forms are diminutive forms of names such as "kedicik" for "kedi" or "anneciğim" for "anne" These forms are not considered as neologisms. However, the data suggests that words such as "aşko" or "güno" in Turkish are hypocoristic words and neologisms. This formation process requires clipping of a word's end

while keeping the head (i.e. “efsane” to “efs”). It then requires the letter -o to be added at the end of the clipped word (i.e. “efs” to “efso”). This process also adds a form of affection and closeness to the new word, with the only exception of “erko”, which is used as a derragotary term. This process may born out of a rare hypocoristic form used on proper names (“Yiğido”, “Hamido”). Another critical finding based on neologism formation is that some formation methods, such as clipping and compounding, initialism, and acronyms, are not found in the corpus. That is to say, they were never the primary methods but were used with borrowings, hybrid neologisms, or hypocoristic neologisms. This may indicate that they are a rarer form of neologism formation for the Turkish language.

Analysing the frequency of the neologism categories also yields important results. It is found that all borrowed neologisms and hybrid neologisms are taken from English. This may be the result of globalisation, the fast spread of information on the internet, and English being the most spoken second language in Turkey.

In terms of coinage, the majority of the neologisms are either newly coined terms or direct borrowings from English. This is further observable in formation processes as they directly correlate with each other.

Function wise, the data suggests expressive neologisms are far more common on social media compared to referential neologisms. This may be due to the informal nature of social media.

Finally, the source of the majority of neologisms found in this study comes from popular culture. A small percentage comes from scientific sources, and no neologisms originating from nonce words or politics were found in the data. This is not surprising, especially for nonce words, as nonce words are not frequent on social media. Considering that the vast majority of social media is used for entertainment, it is not surprising that neologisms originate from popular culture more frequently. Analysing a corpus from a formal source such as newspapers

or books may yield better results for detecting neologisms sourced from science, politics, and nonce words.

Due to limitations, various approaches could not be taken for the study. Gathering a corpus through filter endpoints of X API was not available as a means of compiling the corpus. However, it would yield better results for detecting nonce words and a larger sample size for detecting neologisms. In addition, this resulted in the inability to normalise and study the data in a way that is suited to analysing the geographical distributions of these neologisms.

Another critical approach that can be taken with a corpus gathered with filter endpoints is determining trends and calculating the frequency of a neologism spread throughout the year. Since the corpus for this study was compiled using the sample endpoint, Tweets were gathered from random dates, and the corpus could not get a consistent daily data flow. Analysing these trends on a yearly basis may prove more useful for lexicographers in identifying and studying neologisms for the purposes of dictionary compiling.

The research yields no sociolinguistic results as it is limited to gathering personal data by the X API. Geographical information is the only information relevant to sociolinguistics and can be collected through the API. However, the data collection method does not allow data to be gathered to establish the socioeconomic class, age, gender, or sub-cultures of the users. This requires a more traditional approach with questionnaires, interviews, and direct recordings to be conducted on the participants, similar to Safa & Bilginsoy's study.

The findings of the study show that there are four new neologism formation methods in the Turkish language that were not included in previous studies. These neologism formation methods are hypocoristic neologisms, phono-semantic shift, blending, and hybrid neologisms. Hypocoristic neologisms are suggested as a way of creating neologisms unique to the Turkish language. Hybrid neologisms and borrowings on social media were all taken from the

English language. There were no clippings in the data. In addition, even though compound words are not present in the findings, combined words are considered neologisms in Turkish. This suggests that compounded words written separately are more common in Turkish. The majority of the neologisms found on social media are expressive in function, as expected. In addition, most of the neologisms were sourced from popular culture due to the nature of social media platforms.

BIBLIOGRAPHY

Aksan, D. (1977). *Her Yönüyle Dil: Ana Çizgileriyle Dilbilim*. TDK.

Baayen, R. H., & Renouf, A. (1996). Chronicling the times: Productive lexical innovations in an English newspaper. *Language*, *72*(1), 69. https://doi.org/10.2307/416794

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135-160. https://doi.org/10.1111/josl.12080

Bauer, L. (1983). *English word-formation*. Cambridge University Press.

Bayol, E. M. (2020). *Türkçe İçin Doğal Dil İşlemleme Araçları*. GitHub. https://github.com/brolin59/trnlp

Boulanger, J. (2010). Sur l'existence des concepts de "néologie" et de "néologisme". Propos sur un paradoxe lexical et historique. In M. Teresa Cabré et al. (eds.), Actes del I Congrés Internacional de Neologia de les Llengües Romàniques. Barcelona: Universitat Pompeu Fabra.

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational aspects of retweeting on Twitter. *2010 43rd Hawaii International Conference on System Sciences*, 1-10. https://doi.org/10.1109/HICSS.2010.412

Cabré, M. T. (1993). La terminología. Teoría, metodología, aplicaciones. Barcelona: Empúries.

Cabré, M. T. (2006). La clasificación de neologismos: una tarea compleja. *Alfa 50*(2), 229-250.

Cabré, M. Teresa. (2015). La neologia: un nou camp a la cerca de la seva consolidació científica. *Caplletra 59*, 125-136

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Crystal, D. (2006). *Language and the Internet*. Cambridge University Press.

Crystal, D. (2011). *Internet Linguistics: A Student Guide*. Routledge.

Çelik, M. (1995). *Neologisms in contemporary Turkish*. La Trobe. https://doi.org/10.26181/22201876.v1

Çokol, E. (2020). The language of generation z in the axis of generation conflict. *International Journal of Languages' Education and Teaching*, *8*(2), 404-419

Çağrı Çöltekin. (2014). A set of open source tools for Turkish natural language processing. *Proceedings of the Ninth International Conference on*

Language Resources and Evaluation. European Language Resources Association. https://aclanthology.org/L14-1375/

Dannet, M., & Herring, S. (2007). Linguistic diversity in global communication. *International Journal of Multilingualism, 4*(3), 169-187.

Danet, B., & Herring, S. C. (Eds.). (2007). *The Multilingual Internet: Language, Culture, and Communication Online*. Oxford University Press.

Doğruöz, A. S., & Backus, A. (2009). Innovative constructions in Dutch-Turkish: An assessment of ongoing contact-induced change. *Bilingualism: Language and Cognition*, *12*(1), 41-63.

Eckert, P. (1989). *Jocks and Burnouts: Social Categories and Identity in the High School*. Teachers College Press.

Eisenstein, J. (2013). What to do about bad language on the internet. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics. https://aclanthology.org/N13-1037

Eker, S. (2013). *Çağdaş Türk Dili.* Grafiker Yayınları.

Ergin, M. (1994). *Üniversiteler İçin Türk Dili.* Bayrak Basın Yayım.

Fang, Y.-Q. (2021). A glimpse at the study of english neologism. *Business Prospects*, *2*(2), 78-84. https://doi.org/10.52288/bp.27089851.2021.12.13

Gibson, W. (1984). *Neuromancer*. Ace Books.

Gries, S. T. (2004). Shouldn't it be breakfunch? A quantitative analysis of blend structure in English. *Linguistics, 42*(3), 639-667.

Haugen, E. (1950). The analysis of linguistic borrowing. *Language, 26*(2), 210-231.

Herring, S. C. (1996). Two variants of an electronic message schema. In S. C. Herring (Ed.), *Pragmatics & Beyond New Series* (pp. 81-108). John Benjamins Publishing Company.

Herring, S. C., Stein, D., & Virtanen, T. (2013). *Pragmatics of Computer-Mediated Communication*. De Gruyter Mouton.

Hohenhaus, P. (2005). Lexicalisation and Institutionalization. In P. Štekauer & R. Lieber (Eds.), *Handbook of Word-Formation* (pp. 353-373). Springer.

Huang, Y., Guo, D., Grieve, J., & Kasakoff, A. (2016). Understanding US regional linguistic variation with Twitter data analysis. *Computers, Environment and Urban Systems*, *59*, 244-255.

Işıksal, B. (2015). The sociolinguistic situation in Turkey: An overview. *Journal of Sociolinguistics, 19*(4), 447-468.

Johnstone, B. (1999). Uses of southern-sounding speech by contemporary Texas women. *Journal of Sociolinguistics, 3*(4), 505-522.

Kamola Rafikovna Ismoilova, G. Z. Shamsiddinova, N. A. Maxsudova, & M. Z. Turg'unova. (2023). Neologism in English language and i'ts influence on linguistics. *Journal of Advanced Zoology*, *44*(6), 1790-1793. https://doi.org/10.17762/jaz.v44iS6.2619

Kara, M. (2019). Digital transformations and the evolution of the Turkish language. *Journal of Modern Turkish Studies, 16*(3), 309-328.

Karaağaç. G. (2012). *Türkçenin Dil Bilgisi.* Akçağ Yayınları.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Blackwell.

Lei, S., Yang, R., & Huang, C.-R. (2021). Emergent neologism: A study of an emerging meaning with competing forms based on the first six months of COVID-19. *Lingua*, *258*, 103095. https://doi.org/10.1016/j.lingua.2021.103095

Lewis, G. (2002). *The Turkish Language Reform: A Catastrophic Success*. Oxford University Press.

Llopart-Saumell, E., & Cañete-González, P. (2023). Are stylistic neologisms more neological? A corpus-based study of lexical innovations of women and men. *Languages*, *8*(3), 175. https://doi.org/10.3390/languages8030175

Mace, J. (2013). Memes as tools for promoting linguistic innovation on social media. *Journal of Digital Communication, 15*(2), 47-61.

Mayda, İ., Demir, Y. E., Dalyan, T., & Diri, B. (2021). Hate speech dataset from turkish tweets. *2021 Innovations in Intelligent Systems and Applications Conference*, 1-6. https://doi.org/10.1109/ASYU52992.2021.9599042

Milroy, L. (2002). Social networks. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 549-572). Blackwell.

Milroy, L., & Milroy, J. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics, 21*(2), 339-384.

Özkan, D. (2018). The impact of globalisation on the Turkish language: A study of English loanwords in Turkish business and technology sectors. *Journal of Language and Culture, 7*(1), 23-35.

Oxford Dictionaries. (2013). Oxford Dictionaries Word of the Year 2013: Selfie. Retrieved from https://languages.oup.com/

Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press.

Rodríguez Guerra, A. (2016). Dictionaries of neologisms: A review and proposals for its improvement. *Open Linguistics*, *2*(1). https://doi.org/10.1515/opli-2016-0028

Saussure, F. de (1916). *Course in General Linguistics*. McGraw-Hill.

Shoemark, P. J. (2020). *Discovering and analysing lexical variation in social media text*. https://doi.org/10.7488/era/417

Şafak, Z., & Bilginsoy, M. (2019). Kırklareli merkez örnekleminde Z kuşağı gençlerinin sosyal medyadaki yeni kelimeleri kullanım alışkanlıkları üzerine nicel bir yaklaşım. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, *5*, 125-136. https://doi.org/10.29000/rumelide.606082

Tagliamonte, S. A. (2016). *So Sick or So Cool? The Language of Youth in the Digital Age*. Cambridge University Press.

Tahiroğlu, B.T. (2014). Türkçe Çevrim İçi Haber Metinlerinde Yeni Sözlerin (Neolojizm) Otomatik Çıkarımı. In Özkan, B., Tahiroğlu, B. Tahir ve Özkan Ayşe Eda (Eds.). *Derlem Dilbilim Uygulamaları* (pp. 1-22)ç Karahan Kitabevi Yayınları.

Trudgill, P. (1983). *On Dialect: Social and Geographical Perspectives*. Blackwell.

Wells, J. C. (1982). *Accents of English*. Cambridge University Press.

Yang, G., Hui, P. W., & Li, B. (2012). Hashtag diffusion: A model for the dynamics of user activity on Twitter. *Proceedings of the International Conference on Social Informatics*, 322-335.

Yurtbaşi, M. (2017). Türkiye'de 2000-2017 döneminde oluşan veya yayginlaşan yeni sözler (Neolojizmler). *The Journal of Academic Social Sciences*, *40*(40), 491-562. https://doi.org/10.16992/ASOS.12053

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society, 13*(5), 788-806.

Grieve, J., Nini, A., & Guo, D. (2017). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics, 21*(1), 99–127. doi:10.1017/S1360674316000113

APPENDIX 1. ORIGINALITY FORM

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-YL-15
		Yayın Tarihi *Date of Pub.*	04.12.2023
	FRM-YL-15 Yüksek Lisans Tezi Orijinallik Raporu *Master's Thesis Dissertation Originality Report*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İNGİLİZ DİLBİLİM ANABİLİM DALI BAŞKANLIĞINA

Tarih: 1/11./2024

Tez Başlığı: Sosyal Medya Platformu X'te Türkçe Konuşucuların Kullandıkları Neolojizmlerin Bilgisayar Destekli İncelenmesi
Tez Başlığı (Almanca/Fransızca)*:……

Yukarıda başlığı verilen tezimin a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 117 sayfalık kısmına ilişkin, 1/11/2024 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda işaretlenmiş filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %7 'dır.

Uygulanan filtrelemeler*:

1. ☐ Kabul/Onay ve Bildirim sayfaları hariç
2. ☒ Kaynakça hariç
3. ☐ Alıntılar hariç
4. ☒ Alıntılar dâhil
5. ☒ 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tezimin herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Ad-Soyad/İmza

Öğrenci Bilgileri		
	Ad-Soyad	Berkay Erol
	Öğrenci No	N21137361
	Enstitü Anabilim Dalı	İngiliz Dilbilimi
	Programı	Yüksek Lisans

DANIŞMAN ONAYI

UYGUNDUR.
(Unvan, Ad Soyad, İmza)

* Tez **Almanca** veya **Fransızca** yazılıyor ise bu kısımda tez başlığı **Tez Yazım Dilinde** yazılmalıdır.
**Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları İkinci bölüm madde (4)/3'te de belirtildiği üzere: Kaynakça hariç, Alıntılar hariç/dahil, 5 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit match size to 5 words) filtreleme yapılmalıdır.

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-YL-15
		Yayın Tarihi *Date of Pub.*	04.12.2023
	FRM-YL-15 Yüksek Lisans Tezi Orijinallik Raporu *Master's Thesis Dissertation Originality Report*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

TO HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LINGUISTICS

Date: 1/11./2024

Thesis Title (In English): A Computer Mediated Analysis *of Neologisms Used by Turkish Speakers on X Social Media Platform*

According to the originality report obtained by myself/my thesis advisor by using the Turnitin plagiarism detection software and by applying the filtering options checked below on 1/11/2024 for the total of 117 pages including the a) Title Page, b) Introduction, c) Main Chapters, and d) Conclusion sections of my thesis entitled above, the similarity index of my thesis is 7 %.

Filtering options applied**:

1. ☐ Approval and Decleration sections excluded
2. ☒ References cited excluded
3. ☐ Quotes excluded
4. ☒ Quotes included
5. ☒ Match size up to 5 words excluded

I hereby declare that I have carefully read Hacettepe University Graduate School of Social Sciences Guidelines for Obtaining and Using Thesis Originality Reports that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

Kindly submitted for the necessary actions.

Name-Surname/Signature

Student Information		
	Name-Surname	Berkay Erol
	Student Number	N21137361
	Department	English Linguistics
	Programme	MA

SUPERVISOR'S APPROVAL

APPROVED
(Title, Name and Surname, Signature)

**As mentioned in the second part [article (4)/3]of the Thesis Dissertation Originality Report's Codes of Practice of Hacettepe University Graduate School of Social Sciences, filtering should be done as following: excluding refence, quotation excluded/included, Match size up to 5 words excluded.

APPENDIX 2. ETHICS COMMISSION FORM

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-YL-09
		Yayın Tarihi *Date of Pub.*	22.11.2023
	FRM-YL-09 **Yüksek Lisans Tezi Etik Kurul Muafiyeti Formu** *Ethics Board Form for Master's Thesis*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İNGİLİZ DİLBİLİMİ ANABİLİM DALI BAŞKANLIĞINA

Tarih: 1/11/2024

Tez Başlığı (Türkçe): Sosyal Medya Platformu X'te Türkçe Konuşucuların Kullandıkları Neolojizmlerin Bilgisayar Destekli İncelenmesi Tez Başlığı
(Almanca/Fransızca)*: .....
.....

Yukarıda başlığı verilen tez çalışmam:

1. Insan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırma niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Hacettepe Üniversitesi Etik Kurullarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Ad-Soyad/İmza

Öğrenci Bilgileri		
	Ad-Soyad	Berkay Erol
	Öğrenci No	N21137361
	Enstitü Anabilim Dalı	İngiliz Dilbilimi
	Programı	Yüksek Lisans

DANIŞMAN ONAYI

UYGUNDUR.
(Unvan, Ad Soyad, İmza)

***** Tez **Almanca** veya **Fransızca** yazılıyor ise bu kısımda tez başlığı **Tez Yazım Dilinde** yazılmalıdır.

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-YL-09
		Yayın Tarihi *Date of Pub.*	22.11.2023
	FRM-YL-09 **Yüksek Lisans Tezi Etik Kurul Muafiyeti Formu** *Ethics Board Form for Master's Thesis*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LINGUSTICS

Date: 1/11./2024

ThesisTitle (In English): A Computer Mediated Analysis *of Neologisms Used by Turkish Speakers on X Social Media Platform*

My thesis work with the title given above:

1. Does not perform experimentation on people or animals.
2. Does not necessitate the use of biological material (blood, urine, biological fluids and samples, etc.).
3. Does not involve any interference of the body's integrity.
4. Is not a research conducted with qualitative or quantitative approaches that require data collection from the participants by using techniques such as survey, scale (test), interview, focus group work, observation, experiment, interview.
5. Requires the use of data (books, documents, etc.) obtained from other people and institutions. However, this use will be carried out in accordance with the Personal Information Protection Law to the extent permitted by other persons and institutions.

I hereby declare that I reviewed the Directives of Ethics Boards of Hacettepe University and in regard to these directives it is not necessary to obtain permission from any Ethics Board in order to carry out my thesis study; I accept all legal responsibilities that may arise in any infrigement of the directives and that the information I have given above is correct.

I respectfully submit this for approval.

Name-Surname/Signature

Student Information		
	Name-Surname	Berkay Erol
	Student Number	N21137361
	Department	English Linguistics
	Programme	MA

SUPERVISOR'S APPROVAL

APPROVED
(Title, Name Surname, Signature)