THE CORPUS OF TURKISH YOUTH LANGUAGE (COTY): THE COMPILATION AND INTERACTIONAL DYNAMICS OF A SPOKEN CORPUS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ESRANUR EFEOĞLU ÖZCAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF ENGLISH LANGUAGE TEACHING

SEPTEMBER 2022

Approval of the thesis:

THE CORPUS OF TURKISH YOUTH LANGUAGE (COTY): THE COMPILATION AND INTERACTIONAL DYNAMICS OF A SPOKEN CORPUS

submitted by **ESRANUR EFEOĞLU ÖZCAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in English Language Teaching, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KİRAZCI
Dean
Graduate School of Social Sciences _____

Prof. Dr. Nurten BİRLİK
Head of Department
Department of Foreign Language Education _____

Assoc. Prof. Dr. Hale IŞIK GÜLER
Supervisor
Department of Foreign Language Education _____

Examining Committee Members:

Prof. Dr. Aygül UÇAR (Head of the Examining Committee)
Mersin University
Department of English Linguistics _____

Assoc. Prof. Dr. Hale IŞIK GÜLER (Supervisor)
Middle East Technical University
Department of Foreign Language Education _____

Prof. Dr. Rob DRUMMOND
Manchester Metropolitan University
Department of Languages, Information and Communications _____

Assoc. Prof. Dr. Betil ERÖZ
Middle East Technical University
Department of Foreign Language Education _____

Assoc. Prof. Dr. Turan TEMUR
Anadolu University
Department of Primary Education _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Esranur EFEOĞLU ÖZCAN

Signature:

ABSTRACT

THE CORPUS OF TURKISH YOUTH LANGUAGE (COTY): THE COMPILATION AND INTERACTIONAL DYNAMICS OF A SPOKEN CORPUS

EFEOĞLU ÖZCAN, Esranur

Ph.D., The Department of English Language Teaching

Supervisor: Assoc. Prof. Dr. Hale IŞIK GÜLER

September 2022, 330 pages

This study examines the previously unattained research area of contemporary spoken Turkish used in dyadic and multi-party interaction among young speakers of Turkish. For this purpose, a specialized corpus called the Corpus of Turkish Youth Language (CoTY) was compiled as a source of data and as a tool of analysis. Designed to offer a maximally representative sample of Turkish youth talk, the CoTY contains naturally occurring and spontaneous interactional data among young people between the ages of 14-18 from various socio-economic backgrounds in Turkey. It is a 168,748-word corpus within the single register of informal conversation exclusively among friends. It has 123 unique speakers (62 females and 61 males) and consists of 26 hours 11 minutes of spoken interaction. The corpus was constructed using the multilayer transcription and corpus construction software EXMARaLDA, the tools of Partitur-Editor, COMA, and EXAKT were utilized as corpus building, management, query and analysis tools. The interactional dynamics of the corpus data were examined through four groups of interactional markers; (i) response tokens, (ii) vocatives, (iii) vague expressions, and (iv) intensifiers. For each group of markers; types, distribution, and salient pragmatic functions were presented. The study contributes to sociopragmatic studies of youth language by using systematic, sustainable, and transparent approach to language through corpus methods. It is expected that the results of this study will provide baseline

data for further studies on contemporary spoken Turkish and cross-linguistic youth language studies.

Keywords: corpus linguistics, youth language, spoken discourse, interactional markers, Turkish

ÖZ

TÜRKÇE GENÇLİK DİLİ DERLEMİ (COTY): DERLEM OLUŞTURMA VE SÖZLÜ BİR DERLEMİN ETKİLEŞİMSEL DİNAMİKLERİ

EFEOĞLU ÖZCAN, Esranur
Doktora, İngiliz Dili Öğretimi Bölümü
Tez Yöneticisi: Doç. Dr. Hale IŞIK GÜLER

Eylül 2022, 330 sayfa

Bu çalışma, daha önce odaklanılmamış bir araştırma alanı olan ve genç Türkçe konuşucuları tarafından ikili veya çok taraflı etkileşimde kullanılan çağdaş sözlü Türkçeyi araştırmaktadır. Bu amaçla, veri kaynağı ve analiz aracı olarak Türkçe Gençlik Dili Derlemi (CoTY) adı verilen bir özel alan derlemi oluşturulmuştur. Türkçe gençlik konuşmasının azami düzeyde temsili bir örneğini sunmak üzere tasarlanan CoTY, Türkiye'deki çeşitli sosyo-ekonomik geçmişlerden gelen 14-18 yaş arası gençler arasında plansız ve doğal olarak meydana gelen etkileşimsel sözlü veriyi içermektedir. Bu derlem, sadece arkadaşlar arasındaki gayri resmi konuşmalardan oluşan tek bir dil kesitine ait 168,748 kelimelik bir derlemdir. 123 konuşmacıdan (62 kadın ve 61 erkek) oluşan derlemde, 26 saat 11 dakikalık sözlü etkileşim yer almaktadır. Derlem, çok katmanlı transkripsiyon ve derlem oluşturma yazılımı EXMARaLDA kullanılarak oluşturulmuş; Partitur-Editor, COMA ve EXAKT araçları, derlem oluşturma, yönetim, sorgulama ve analiz araçları olarak kullanılmıştır. Derlem verilerinin etkileşimsel dinamikleri dört grup etkileşim belirleyicisi üzerinden incelenmiştir: (i) yansıma belirteçleri, (ii) hitap sözcükleri, (iii) belirsizlik ifadeleri ve (iv) pekiştiriciler. Her bir etkileşim belirleyicisi grubu için; türler, dağılım ve göze çarpan edimsel işlevler sunulmuştur. Çalışma, derlem yöntemleri aracılığıyla dilbilim çalışmalarına sistematik, sürdürülebilir ve şeffaf bir yaklaşım sunmakta ve gençlik dilinin sosyopragmatik

incelemelerine katkıda bulunmaktadır. Bu çalışmanın sonuçlarının, çağdaş konuşma Türkçesi ve diller arası gençlik dili çalışmaları için temel veri sağlaması beklenmektedir.

Anahtar Kelimeler: derlem dilbilim, gençlik dili, sözlü söylem, etkileşim belirleyicileri, Türkçe

to my grandfather

ACKNOWLEDGMENTS

This dissertation owes a big gratitude for the support, guidance, and patience of a great many people.

I am grateful to be able to complete this study under the supervision of Assoc. Prof. Dr. Hale Işık-Güler who has been not only an academic advisor but a life mentor to me for over ten years. Her presence has always encouraged me to push my capabilities as a researcher and pursue innovation in linguistic research.

I would like to thank my committee members Prof. Dr. Aygül Uçar, Assoc. Prof. Dr. Betil Eröz, Assoc. Prof. Dr. Turan Temur, and Prof. Dr. Rob Drummond for their insights and immense encouragement for this work. I would also like to recognize Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting my doctoral studies.

I wish to extend my appreciation to an outstanding scholar who has been influential on my academic journey: Prof. Dr. Şükriye Ruhi. I remain grateful to her for introducing me to corpus linguistics and giving me the opportunity to contribute to the Spoken Turkish Corpus when I was an undergraduate student. Her thought-provoking questions about spoken discourse apparently shaped my path without me even realizing it back then.

Many amazing people have contributed to this work in various ways. I'd like to thank the loveliest academic support circle one can ever have, Merve Bozbıyık and Yasemin Erdoğan-Öztürk, for reminding me to breathe every now and then, also to recognize and celebrate the milestones in our doctoral endeavours as young researchers; my dearest friend Banu Çiçek Başaran-Uysal for always being there to listen to me, guide me, and uplift me; Ayşenur Erdemir for accompanying me in our never-ending quest of cultivating the mindfulness within academia; my friends and fellow doctoral students Melike Demir-Bektaş, Özlem Özbakış and Gizem Doğan for rooting for me all the time; the DISCORE core members for fruitful discussions and data sessions; Joshua Bear and Deniz Şallı-Çopur for being such exceptional inspirations for me since my undergraduate

years at METU; Yiğithan Bektaş, all the gatekeepers and participants who responded to the call for this project and helped me with the data collection; and my colleagues at the Council of Higher Education for their continual support.

Warmest thanks go to my family. I am indebted to my parents Nevin-İhsan Efeoğlu, brothers Hulusi Eren and Alperen for the endless love, devotion, and scholarly vision; and I am extremely lucky to have my Özcan family, Afet-Yusuf Özler Özcan, Ceyhun Özcan, and Buse Özcan, as a source of affection, kindness and support in my life.

Last but the greatest thanks is for my husband, Cem, for his love, company, and humour. Thank you for the strength you have given me to complete this dissertation, cheering me up and keeping me grounded throughout the ups and downs of this journey.

Thank you all.

TABLE OF CONTENTS

PLAGIARISM iii

ABSTRACT iv

ÖZ vi

DEDICATION viii

ACKNOWLEDGMENTS ix

TABLE OF CONTENTS xi

LIST OF TABLES xvii

LIST OF FIGURES xix

LIST OF ABBREVIATIONS xx

CHAPTERS

1. INTRODUCTION 1

1.0 Presentation 1

1.1 Background to the Study 1

1.2 Problem 4

1.3 Purpose and Scope 6

1.4 Significance of the Study 7

1.5 Limitations 8

1.6 Organization of the Dissertation 10

2. REVIEW OF LITERATURE 12

2.0 Presentation 12

2.1 Defining youth and youth language 12

2.2 Research methods in youth language 15

2.2.1 Variationist studies 16

2.2.2 Corpus driven studies 17

2.2.2.1 English youth talk corpora 18
2.2.2.2 Spanish youth talk corpora 20
2.2.2.3 German youth talk corpora 22
2.2.2.4 Nordic youth talk corpora 23
2.2.2.5 Spoken learner corpora 24
2.3 Recent foci of investigation 26
2.3.1 Indexing identities 27
2.3.2 Linguistic innovation and change 28
2.3.3 Multilingual encounters 30
2.3.4 Stylization in digital sphere 31
2.4 Youth studies in Turkey 33
2.4.1 Informing fields 33
2.4.2 Linguistics 36
3. METHOD OF RESEARCH 40
3.0 Presentation 40
3.1 Research Design 40
3.1.1 Research Questions 41
3.1.2 Authenticity 43
3.1.3 Representativeness 44
3.1.4 Size 47
3.2 Participants 48
3.2.1 The Sampling Frame 48
3.2.2 Recruiting Participants 49
3.2.3 Roles of Participants 50
3.3 Data Sources and Data Collection Timeframe 51
3.3.1 Interviews 52
3.3.2 Audio Data 52
3.3.3 Recording Log 54

3.4 Piloting 54
3.5 Corpus Building and Analysis Software Used: EXMARaLDA 58
3.5.1 Partitur-Editor 60
3.5.2 COMA 60
3.5.3 EXAKT 61
3.5.4 Workflow 62
3.6 Scope of Metadata in the CoTY 63
3.6.1 Communication Metadata 65
3.6.2 Speaker Metadata 66
3.7 Transcription and Annotation 67
3.8 Method of Analysis 72
3.8.1 Corpus-oriented discourse studies 72
3.8.2 Corpus analytical methods 73
3.8.2.1 Frequency lists 74
3.8.2.2 KWIC analysis 74
3.8.2.3 Co-occurence 75
3.8.2.4 Interpreting Discourse 76
3.9 Reliability and Validity 76
3.10 Ethical Considerations 77
4. FINDINGS AND DISCUSSION 79
4.0 Presentation 79
4.1 Corpus structure 79
4.1.1 Scope 80
4.1.2 Speakers 80
4.1.3 Domain 88
4.2 Topical and lexical characteristics 93
4.2.1 Topics 93
4.2.2 Key concepts and typical vocabulary 103

4.3 Interactional markers 107
4.3.1 Response tokens 108
4.3.1.1 Defining response tokens 108
4.3.1.2 Brief overview of related work on response tokens 109
4.3.1.2.1 Response tokens in Turkish 110
4.3.1.2.2 Response tokens in youth talk 111
4.3.1.3 Findings: Response tokens in the CoTY 112
4.3.1.3.1 Non-lexical response tokens 114
4.3.1.3.2 Lexical response tokens 119
4.3.1.3.3 Unconventional forms: Words of slang, taboo, religion 125
4.3.1.3.4 Register-specificity: *aynen* 129
4.3.2 Vocatives 140
4.3.2.1 Defining vocatives 140
4.3.2.2 Brief overview of related work on vocatives 142
4.3.2.2.1 Vocatives in Turkish 142
4.3.2.2.2 Vocatives in youth talk 144
4.3.2.3 Findings: Vocatives in the CoTY 147
4.3.2.3.1 Types, addressers, and addressees 147
4.3.2.3.2 Forms and positions 150
4.3.2.3.3 Semantic categories 152
4.3.2.3.4 Referents 159
4.3.2.3.5 Multiplicity of functions: *kanka* 160
4.3.2.3.6 Address shifts: *bro* versus *kardeş* 165
4.3.3 Vague expressions 168
4.3.3.1 Defining vague expressions 169
4.3.3.2 Brief overview of related work on vague expressions 171
4.3.1.3.1 Vague expressions in Turkish 171
4.3.1.3.2 Vague expressions in youth talk 172

4.3.3.3 Findings: Vague expressions in the CoTY 174
4.3.3.3.1 Types and distribution 174
4.3.3.3.2 Vague references 176
4.3.3.3.3 Vague additives 181
4.3.3.3.4 Communicative purposes and vague language: *f(a)lan* 189
4.3.4 Intensifiers 208
4.3.4.1 Defining intensifiers 209
4.3.4.2 Brief overview of related work on intensifiers 210
4.3.4.2.1 Intensifiers in Turkish 211
4.3.4.2.2 Intensifiers in youth talk 211
4.3.4.3 Findings: Intensifiers in the CoTY 213
4.3.3.3.1 Types, distribution, and speakers 214
4.3.4.3.2 Amplifiers 219
4.3.4.3.3 Downtoners 222
4.3.4.3.4 Tracing delexicalization: From *tık* to *bi tık* 224
4.3.4.3.5 Accentuating the personality traits: *aşırı* or *bayağı* 229
5. CONCLUSION 233
5.0 Presentation 233
5.1 Summary of Findings 233
5.1.1 Layer One: Corpus Construction 233
5.1.2 Layer Two: Linguistic Architecture 237
5.2 Future Directions 246
REFERENCES 249
APPENDICES
A. ONLINE FLYER 291
B. INFORMED CONSENT FORMS 292
C. DATA COLLECTION TIMELINE 294
D. SEMI-STRUCTURED INTERVIEW PROTOCOL 296

E. RECORDING LOG....297
F. TRANSCRIPTION CONVENTIONS....300
G. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE....302
H. THE PARTICIPANT SAMPLE....303
I. VOCATIVES WITH GENDERED REFERENTS....306
J. CHI-SQUARE TEST RESULTS....308
K. CONCORDANCES FOR AŞIRI AND BAYAĞI....309
L. VITA....310
M. TURKISH SUMMARY / TÜRKÇE ÖZET....313
N. THESIS PERMISSION FORM / TEZ İZİN FORMU....330

LIST OF TABLES

Table 1 Register characteristics of the CoTY 45
Table 2 Youth population by province and sex in Turkey 46
Table 3 Pilot study participants 55
Table 4 Comparison of multilayer transcription software programmes 58
Table 5 Comparison of metadata in existing spoken youth language corpora 64
Table 6 Transcription Conventions (HIAT) 70
Table 7 Paralinguistic and prosodic features annotated in the corpus 71
Table 8 Distribution of grade levels and age ranges tabulated by speaker sex 81
Table 9 Structure of types of conversations in the CoTY 83
Table 10 Youth population in Istanbul, Ankara and Izmir versus the CoTY 85
Table 11 Description and distribution of socio-economic status levels in the CoTY 86
Table 12 Types of interaction and duration of recordings tabulated by pandemic related events during data collection period 90
Table 13 Type of interaction, setting, and ongoing activities in the CoTY 92
Table 14 Distribution of topics and sub-topics 94
Table 15 Keywords in the domain of daily life and education 105
Table 16 Keywords in the domain of interactional markers 106
Table 17 Non-lexical response tokens tabulated by frequencies and speakers 114
Table 18 Collocates and reduplications of non-lexical response tokens 118
Table 19 Lexical response tokens tabulated by frequencies and speakers 120
Table 20 Collocates and reduplications of lexical response tokens 124
Table 21 Distribution of *aynen* by speaker age 131
Table 22 Distribution of *aynen* tabulated according to speaker groups 132
Table 23 Functional distribution of *aynen* 133
Table 24 Distribution of vocatives in the CoTY 148
Table 25 Distribution of positions of vocatives in utterances 151
Table 26 Semantic categories of vocatives 152
Table 27 Functions of *kanka* 161
Table 28 Comparison of functions of *bro* versus *kardeş* as vocatives 166
Table 29 Distribution of vague expressions 175

Table 30 Vague reference tokens tabulated by frequencies and speakers.....176
Table 31 Vague additive tokens tabulated by frequencies and speakers.....182
Table 32 Distribution of *f(a)lan* by speaker age.....190
Table 33 Overview of communicative purposes of DUs in which *f(a)lan* is used.....194
Table 34 Amplifiers and downtoners tabulated by frequencies and speakers.....215
Table 35 The most frequent 10 intensifiers for females and males.....218
Table 36 Profiles of spoken corpora for youth language.....236

LIST OF FIGURES

Figure 1 Foci of investigation in each research question layer in the study..... 42
Figure 2 Screenshot of Partitur-Editor interface..... 60
Figure 3 Screenshot of COMA interface..... 61
Figure 4 Screenshot of EXAKT interface..... 62
Figure 5 Corpus construction workflow..... 63
Figure 6 Distribution of speaker ages at the time of recording tabulated by tokens..... 81
Figure 7 Distribution of tokens by sex of speakers and types of speaker groups..... 82
Figure 8 Hometowns of speakers in the CoTY..... 84
Figure 9 Cities of residence of speakers in the CoTY..... 84
Figure 10 Distribution of tokens with regard to socio-economic status of speakers..... 87
Figure 11 Frequency of communication between speakers within groups..... 89
Figure 12 100 most frequently occurring tokens in the CoTY versus the STC..... 104
Figure 13 Sample concordance lines for *aynen* in the CoTY..... 130
Figure 14 Sample concordance lines for *kanka* in the CoTY..... 160
Figure 15 Sample concordance lines for *f(a)lan* in the CoTY..... 190
Figure 16 Sample concordance lines for *bir tık* in the CoTY..... 225

LIST OF ABBREVIATIONS

BNC	British National Corpus
BNC2014S	British National Corpus 2014 Spoken
CANCODE	Cambridge and Nottingham Corpus of Discourse in English
COLA	Corpus Oral de Lenguaje Adolescente
COLSEC	College Learners' Spoken English Corpus
COLT	Bergen Corpus of London Teenage Language
CORMA	El Corpus Oral de Madrid
CoTY	Corpus of Turkish Youth Language
CYLIL	Corpus of Young Learner Interlanguage
DCPSE	The Diachronic Corpus of Present Day Spoken English
EU	European Union
EVA	Evaluation of English in Norwegian Schools
EXMARaLDA	Extensible Mark-up Language for Discourse Annotation
HIAT	Semi-Interpretative Working Transcriptions
ICE	International Corpus of English
ICLE	International Corpus of Learner English
ISLC	The Icelandic Spoken Language Corpus
JuBE	Jugendsprache Schweiz Korpus
KidKo	The KiezDeutsch-Korpus
KWIC	Key Word In Context
LCIE	Limerick Corpus of Irish English
LCSWE	Longman Corpus of Spoken and Written English
LINDSEI	Louvain International Database of Spoken English
MLE	The Multicultural London English Corpus
MONE	Turkish Ministry of National Education
MTC	METU Turkish Corpus
NEET	Not in Employment, Education, or Training
NESSI	New Englishes Student Interviews Corpus
Ph@ttSessionz	Ph@ttSessionz Speech Database

	System Aided Compilation and Open Distribution of European
SACODEYL	Youth Language
SCoSE	Saarbrücken Corpus of Spoken English
SES	Socio-Economic Status
STC	Spoken Turkish Corpus
TCEEE	Tübingen Corpus of Eastern European English
TDK	Turkish Language Association
TIC	Toronto Internet Corpus
TIMC	Toronto Instant Messaging Corpus
TNC	Turkish National Corpus
TTC	Toronto Teen Corpus
TUİK	Turkish Institute of Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNO	Språkkontakt och ungdomsspråk i Norden
UrBEn-ID	Urban British English and Identity Project
VOYS	Voices of Young Scots
WCSNZE	Wellington Corpus of Spoken New Zealand English
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

« La jeunesse n'est qu'un mot. »
Youth is just a word.

Pierre Bourdieu, 1978

1.0 Presentation

This introductory chapter presents the study by giving information on the background to the dissertation, the problem this study aims to touch upon, the purpose and the scope of the research conducted, and the significance of the study. Lastly, the limitations are presented and explained.

1.1 Background to the Study

Over course of evolving agendas adopted for investigating linguistic variation and social meaning, the *age* of speakers has been used as a parameter for depicting the boundaries of different speech groups in a community and explore their shared as well as divergent linguistic practices with regard to their greater community. *Youth language* has often attracted the attention of researchers due to its dynamic, fluid, and performative nature. By investigating linguistic behaviour of younger speakers, researchers have been exploring not only the contemporary account of a language but also the trajectories of language change.

There is no single definition of youth language as *the youth* itself is a fuzzy and socially constructed category (see Chapter Two for an overview). As a result, youth language practices are by no means homogenous (Martínez, 2011), rather they cover intertwined

facets of linguistic patterns and socio-pragmatics strategies manipulated by speakers in online and offline interaction. These linguistic patterns and strategies have been explored via various modes of data sources and methodological approaches so far.

While the first-wave sociolinguistic studies mainly made use of elicited data and generalized the findings based on static socio-demographic categories; the recent sociolinguistic work utilizes naturally occurring data and approaches the linguistic practices within a new agenda of performative views of language. In such studies, researchers who examine the relationship between the variation in the patterns of linguistic practices and social constructs have dwelled upon the concept of *discourse.* Among its various definitions, discourse can be defined as the linguistic practices in a particular community in which there is a reciprocal relationship between distinct social and situational contexts and language use (Jaworski & Coupland, 2006; Paltridge, 2011).

Compared to written forms of discourse, *spoken discourse* exhibits distinct features such as its fast-changing nature, embedded pragmatic functions, and fragmented structure (Cutting, 2011). To explore these characteristics, the social variables in interaction stand out as vital. These social variables are often challenging to handle as they are intricate and sometimes fuzzy to identify consistently. Rampton (2006), for instance, highlights that boundaries of social categories are now less clear and thus the focus of interactional analyses should be on *the role that language plays* when the categories such as group membership, age, ethnicity contribute in some way to the interaction. This stance on language treats speakers as active agents manipulating the language by deploying linguistic and semiotic resources to accomplish various pragmatic goals. This view is also a reflection of Butler's (1990) work on *performativity* which has had tremendous insights for the study of language and social meaning. The performative turn in linguistics emphasized the negotiation of identities and experimentation with styles within dynamic discourses jointly constructed by interactants.

While the performative turn in linguistics embraces the fluidity of categorizations and discursive meaning making practices, it also led to discussions concerning ensuring the rigour and systematicity across linguistic research (Berez-Kroeker, 2017; Greckhamer, & Cilesiz, 2014). The rise of *open science initiative,* at this point, provides a promising roadmap for the future of language studies. The umbrella term *open science* refers to idea that scientific knowledge -where appropriate- should be accessible, rigorous,

reproducible, replicable, accumulative, inclusive (Abele-Brehm et al., 2019; Kathawalla et al., 2020; Syed, 2019; Woelfe et al., 2011). Within this line, the initiative calls for the implementation of transparent and collaborative approaches to knowledge creation and dissemination (Fecher & Friesike 2014). *Corpus linguistics,* as an answer to this call, offers a relatively less obtrusive method for data collection, a sustainable tool to conduct multiple layers of linguistic queries for research agenda, and a more robust system for the systematic inquiry of a language.

A *corpus* (pl. corpora) is defined as a large body of linguistic evidence composed of attested language use (McEnery, 2005, 2012). Corpora can take various forms in accordance with the purposes they are designed to serve or the characteristics of the linguistic data they have. Corpora can be classified based on their modalities, namely *written*, *spoken*, *multimodal*, or a combination of these. *Monolingual corpora* represent a single language while *parallel corpora* enable researchers to compare the forms of translation for the same text in two languages (e.g., English-Swedish Parallel Corpus) and *comparable corpora* show original texts in two or more languages with same sampling frame (e.g., The English Comparable Corpus) for comparable linguistic analyses. Size can be another classification; a corpus can be built to represent an entire language/variety and thus be labelled as a *general* corpus (e.g., British National Corpus), or it can be a *specialized* corpus designed to represent a language within the boundaries of limited subject areas, genres, domains or topics (e.g., Michigan Corpus of Academic Spoken English). Other forms of possible classifications include *historical corpora* which cover data of different periods of same language (e.g., Helsinki Corpus of English), *monitor corpora* which aim to track language change as it is constantly updated and thus grows over time, (e.g., Corpus of Contemporary American English) *learner corpora* which consist of language learners' spoken and written linguistic output and utilized for pedagogical purposes (e.g., International Corpus of Learner English), and *developmental corpora* which provide evidence for different developmental stages of first language acquisition process (e.g., CHILDES corpus).

Compared to written corpora available across languages, the number of spoken corpora is few particularly due to their challenges and costs. The number of specialized spoken corpora focusing on youth talk is even more limited. Though limited in terms of number (see Chapter Two for details), the available youth talk corpora such as The Bergen Corpus of London Teenage Language (COLT) and Corpus Oral de Lenguaje Adolescente

(COLA) have proved extensive opportunities to pinpoint typical features of the language used by a specified age cohort in a robust and systematic way. As an under-researched topic of investigation of an under-represented language, the defining linguistic characteristics of Turkish youth interaction have been *invisible* within both Turkish linguistics and cross-linguistic studies so far.

With the development of corpus linguistics as a methodological approach to the language, a consistent and reliable approach to the study of linguistic patterns in relation to their situational and social variables. In this line, this study incorporates corpus linguistics into the study of contemporary Turkish spoken by Turkish youth and adheres to open science practices to contribute to the growing body of consistent, sustainable, accountable investigation in linguistic. Through the compilation of first corpus of youth language for Turkish, namely the Corpus of Turkish Youth Language (CoTY), and employing corpus linguistic tools to the systematic study of the authentic language data, this study presents a baseline investigation to examine the multiple interactional facets of youth interaction.

1.2 Problem

Even though there is a substantial body of work focusing on the linguistic practices of the youth in several languages such as English, Spanish, and German among others; the studies in Turkish are scarce. The majority of the existing studies do not offer rigorous analyses of the issue but rather offer a relatively deterministic perspective on the linguistic practices of Turkish youth with limited or no interactional data. As for corpus-based studies, no study has integrated corpus linguistics tools to investigate the interaction among Turkish youth yet and there is no a specialized corpus focusing on Turkish youth talk either.

So far, there have been three prominent corpus construction initiatives, namely Middle East Technical University Turkish Corpus[1], Turkish National Corpus[2] and The Spoken Turkish Corpus[3], within the field of corpus linguistics in Turkey.

The first linguistic corpus to represent contemporary Turkish is The Middle East Technical University Turkish Corpus (MTC) which is a 2-million-word written corpus. It consists of data from 1990-2002 in 10 different genres and it is designed to be a balanced corpus (Say et al., 2004). There are two sub-corpora of this corpus, namely The METU-Sabancı Turkish Treebank (Oflazer et al., 2003) and METU-Turkish Discourse Bank Project (Zeyrek et al., 2013). The METU-Sabancı Turkish Treebank is morphologically and syntactically annotated sub-corpus of 65,000-words while the METU-Turkish Discourse Bank Project focuses on discourse annotation for 400,000-word sub-corpus of the MTC. As the corpora was collected through opportunistic sampling, some genres are more overrepresented than others, and the corpus solely depends on written data. Still, the MTC and its sub-corpora represent outstanding works as the predecessors of Turkish corpora.

Turkish National Corpus (TNC) was designed as general corpora of contemporary Turkish. Built at Mersin University (Aksan, Aksan, Koltuksuz, Sezer, Mersinli, Demirhan, Yılmazer, Atasoy, Öz, Yıldız, & Kurtoğlu, 2012), the large-scale Turkish National Corpus (TNC) consists of 50,000,000 words, the majority of which are drawn from written texts (98%). Transcribed spoken data (2%) constitutes the remaining portion of the corpus. The written part includes a wide range of genres covering a time span from 1990 to 2013 (24 years), and the spoken portion comprises of spontaneous, every day conversations collected in particular communicative settings. The corpus has morphological and part-of-speech annotation and provides an online query interface available for research purposes. A wide range of query options are available for the spoken portion such as the gender, education level, socioeconomic background (high, middle, low), general activity and interaction type (comprised of monologues and dialogues). However, it is not

[1] https://ii.metu.edu.tr/metu-corpora-research-group for more information about this corpora project.

[2] https://v3.tnc.org.tr/tnc/about-tnc for more information about this corpus.

[3] https://std.metu.edu.tr/en/ for more information about this corpus.

possible to run queries by speaker age. There is no information provided regarding the age range of the speakers for the totality of the spoken corpus, either.

The only solely spoken corpus of Turkish, The Spoken Turkish Corpus (STC) was constructed at Middle East Technical University (Ruhi, Hatipoğlu, Eröz-Tuğa, Işık-Güler, Acar, Eryılmaz, Can, Karakaş, & Çokal-Karadaş, 2010). The STC is comprised of face-to-face or mediated interactions that were recorded between the years 2008 and 2013 in various regions of Turkey. A demo version is publicly available. The second, beta version of the STC, which is available in-house at METU consists of 50 hours of recording and 350,000 words. The corpus has morphological and pragmatic (speech act) annotation, the transcriptions are presented with their time-aligned audios. The STC offers a highly rich metadata to enable researchers to explore the corpus socio-pragmatically. As it was designed and constructed to represent general spoken Turkish spoken by adult speakers, the overview of speaker ages reveal that the corpus is not able to represent youth talk as it contains only 10 speakers between the ages 10-19.

Due to the fact that all three pioneering corpora[4] were constructed to represent general written and/or spoken contemporary Turkish, neither of them allows for an in-depth description and analysis of youth talk in Turkey. To fill this gap, this study stands out as the presentation of the first spoken corpus of Turkish youth language.

1.3 Purpose and Scope

This study explores the previously unexplored research area of contemporary spoken Turkish spoken by the youth through two complementary goals reflected via the research questions (see section 3.1.1 of Chapter Three for research questions). Firstly, the study aims to develop *the tool*, a specialized corpus, to enable the sustainable investigation of Turkish and cross-linguistic youth talk. Secondly, the study aims to *use this tool* to examine the linguistic dynamics of talk in terms of its macro structures such as topics and micro structures such as interactional markers in this dyadic and multi-party interaction.

[4] In addition to these university-affiliated and pioneering Turkish corpora projects, there are also a range of independent corpora projects such as TS Corpus (Sezer & Sezer, 2013) which is a large collection of corpora compiled from web sources such as online newspapers, forums, blogs, etc. Please also see Çöltekin et al. (2022) for a comprehensive survey of other corpora and lexical resources available for Turkish.

In line with these aims, there are two sub-goals behind constructing the first spoken corpus of Turkish youth language. Firstly, the corpus aims to contribute to the growing studies in *corpus linguistics and corpus methodology in Turkey.* Secondly, the corpus aims to provide *a cross-linguistic perspective* for the existing literature on youth language studies which so far have focused on research based on English and Spanish, and to some extent; German and Nordic languages.

The first goal involves a meticulous and labour-intensive corpus construction process which will be presented in detail in Chapter Three. In order to contribute to the spoken corpus construction methodology, this study aims to present a roadmap for future corpora design by presenting the criteria and justifications adopted for the design, data collection, transcription, annotation stages of the corpus.

The second goal is related to exploring the linguistic characteristics of the interaction within the corpus which consists of topics, sub-topics, key concepts and keywords. In terms of interactional characteristics of the data within the CoTY, the scope of investigation focuses on four main categories of linguistic entities within the corpus, namely (i) *response tokens*, (ii) *vocatives,* (iii) *vague expressions*, and (iv) *intensifiers* which were selected based on the results of the keyness analysis conducted between the CoTY and the Spoken Turkish Corpus.

1.4 Significance of the Study

The past thirty years has witnessed the rise of corpora as both tools and sources of data for linguistic investigations. Among numerous corpora projects, specialized spoken corpora are still few in number due to the overwhelming amount of time, human resources, and funding required to compile and build them compared to written corpora projects. Among them, the number of spoken youth talk corpora are even more limited.

To fill this gap, this study presents the compilation of the first spoken corpus of youth talk in an under-represented language, Turkish. The most fundamental contribution of the CoTY is that it provides the baseline data to examine linguistic and relational dynamics of youth talk which was not available for Turkish until now. By examining the most salient features of the corpus, the study aims to present a solid ground for the future

investigations regarding both cross-linguistic youth language research and Turkish linguistics.

The majority of studies in Turkish linguistics are based on written data. Through the construction of the CoTY, a rich and sustainable resource of naturally occurring data is generated as a complementary perspective to the scholarly knowledge accumulated so far. Furthermore, constructing a specialized corpus which is designed to expand over time enables monitoring the changes in both Turkish youth language and spoken Turkish, and also lays the groundwork for future corpus studies adopting diachronic perspectives to language research.

An additional facet of significance of this study is its overarching design advocating *open science* practices in linguistics. By utilizing the affordances of corpus tools to sustain reproducibility, consistency, and transparency in language research, this study promotes the open science initiative. The conscious decisions made by the researcher regarding the utilization of contributory public participation model (Shirk, et al., 2012) to integrate public engagement; the use of an open source corpus construction and annotation software EXMARaLDA to ensure the sustainable development of the corpus in terms of size and levels of annotation in the future; providing access to the schemes for conventions, annotation, and metadata adapted or developed for the corpus in an open access repository[5] also resonate with the aims of the open science movement.

1.5 Limitations

As with all forms of research, the study at hand bears a number of limitations. Spoken corpus compiling and construction is a time-consuming and resource-intensive process, therefore several compromises were made to adopt a feasible, yet valid approach to complete this project. Due to the fact that the current study is a dissertation study with a single researcher working within an allocated time frame and with no project funds, the sampling frame and scope of investigation were designed to be practical enough for the researcher to handle the data and robust enough to ensure reliability and validity of the study.

[5] The repository for the Corpus of Turkish Youth Language can be accessed via https://osf.io/ek4z8/

First and foremost, the CoTY is a specialized corpus and it should be noted that it is not representative of the entire youth population in Turkey. As will be presented in detail in Chapter Three, a maximally representative sample was obtained in order to construct the corpus. The results of the analyses, then, are not generalizable to the greater population yet they hint implications for overall dynamics. For instance, even though the study collected the data of young speakers between ages 14 to 18 across the country, the sample does not include NEETs[6] and 14-18 year-olds who are actively in labour. As a result, the profile of young speakers provides a partial reflection of the language spoken by Turkish youth. To address this limitation, the study underlines that the participants are young, high schooler speakers of Turkish who are in formal full-time education in Turkey. The community of young people who are neither in education nor in employment or training is composed of vulnerable and often marginalized group in the society, thus reaching out to this group needed a separate approach in terms of recruiting the participants, data collection which was not within the scope of this dissertation study.

Additionally, in terms of its sampling frame, the researcher set the number of interactants in a group to maximum three people and briefed the participants to do recordings accordingly. As a result, the multi-party interaction was limited to three speakers for this corpus project. While this may be viewed as a limitation, the primary justification behind this decision was based on the existing literature on corpus construction project reports and the results of pilot study indicating that speaker identification and decoding of the overlaps in speaker turns are infeasibly time-consuming when there are more speakers. Given the limitations of time and the human resources, the CoTY included only conversations among two or three speakers.

Another inherent limitation is related to the metadata compiled. As it is the case with all kinds of corpora projects, metadata regarding the speaker demographics were dependent on the self-statements of the participants and the informants. For instance, socioeconomic status of the participants in the CoTY were retrospectively coded based on the information regarding the occupations and the education levels of parents of the speakers only. In order to obtain as reliable data as possible, the recording log was

[6] OECD (2022), Youth not in employment, education or training (NEET) (indicator). doi: 10.1787/72d1033a-en (Accessed on August 2022)

designed and formatted in the most cognitively less demanding and less time-consuming format as possible.

In terms of balance and representativeness of the corpus, the sampling frame was not designed to control the distribution of data across speaker sex and types of speaker groups. In order to attain a maximally representative sample, the researcher only focused on reaching a balanced ratio of sex of speakers in the whole corpus. As a result, while the number of female and male speakers is balanced in the corpus; the distribution of spoken data across the speaker sex and speaker groups are skewed. It should be noted, though, that this distortion was the inherent consequence of unobtrusive data collection measures as the speakers were briefed to talk naturally and without any time limit, thus, the length of talk varied for each speaker and speaker group.

The major focus of linguistic analyses conducted using the CoTY was interactional markers under which four groups of linguistic entities which are *response tokens, vocatives, vague expressions* and *vocatives* were examined. While the corpus provides numerous other possibilities for research foci, these categories were chosen based on the results of the keyness analysis so that the salient characteristics of the interaction in the corpus can be presented as the first step of laying ground for future corpus driven studies of Turkish youth talk.

Lastly, as a result of methodological constraints, the current version of CoTY only was orthographically transcribed, lemmatized and pragmatically annotated. Due to the fact that a POS-tagger for Turkish is not integrated into the corpus construction software EXMARaLDA, the corpus does not support any Part-of-speech tagging which is a constraint in terms of defining the scope of analysis that can be conducted and the amount of time required to carry out linguistic analyses on the corpus data for now.

1.6 Organization of the Dissertation

This dissertation consists of five chapters. **Chapter One** introduces the background, the problem, the purpose and scope, the significance, and the limitations of the study. **Chapter Two** reviews the related literature with regard to conceptualizations of the youth and youth language, major research methods utilized to study youth language (i.e., variationist studies and corpus driven studies), the available corpora across languages

built for youth talk (i.e., English, Spanish, German, and Nordic youth talk corpora) as well as spoken learner corpora, recent research themes in youth language (i.e., indexing identities, linguistic innovation an change, multilingual encounters, and stylization in digital sphere), and finally an overview of youth studies in Turkey with a complementary account of research in linguistics and other informing fields. **Chapter Three** presents the corpus compilation of the Corpus of Turkish Youth Language and construction methodology adopted. It starts with presenting the research design, comparison of existing corpora in Turkish, the three pillars of a corpus (i.e., authenticity, representativeness, and size), the workflow of corpus construction using EXMARaLDA, the detailed information on participants in the project, the data sources and the timeframe of the corpus, the scope of metadata, transcription and annotation stages. Later, the corpus analytical methods and corpus approaches to discourse analysis are presented as the methods of analysis. Finally the issues of reliability, validity and ethical considerations are presented and discussed. **Chapter Four** presents the analysis conducted on the corpus. Firstly, the structure of the corpus in terms of its size, speakers, types and tokens is depicted. Later, the main topics identified and the interactional markers focused are presented. The interactional markers are described, presented and discussed under four sub-chapters, namely (i) response tokens, (ii) vocatives, (iii) vague expressions, and (iv) intensifiers. For each sub-chapter, the terminology is defined, a brief overview of related literature is outlined, and the findings are presented along with excerpts from the corpus. **Chapter Five** concludes the dissertation by providing a summary of profile of the constructed corpus, the findings, and the implications for future corpus construction and youth language research.

CHAPTER 2

REVIEW OF LITERATURE

2.0 Presentation

In this chapter, an account of literature related to youth language and corpus linguistics will be presented. Firstly, two of the prominent research approaches to studying youth language, namely variationist studies and corpus driven studies will be outlined. Corpus driven studies will be presented in detail along with the major spoken youth language corpora in English, Spanish, German, and Nordic languages. Additionally, spoken learner corpora will be mentioned. Next, a selection of the recent foci of linguistic investigation carried out within the scope of youth language will be provided. Later, youth studies in Turkey will be outlined by presenting a brief account of research in informing fields in Turkey and finally the review of linguistic research on Turkish youth language will be provided.

2.1 Defining youth and youth language

The concept of *youth* has demonstrated shifting denotations in different cultural and political settings over the course of history. There have been different labels such as adolescents, youth, teenagers which are sometimes used interchangeably without any clear definitions. Different institutions provide different age ranges for the people defined as the youth, such as the categorizations of 15-24 years for UNESCO, 10-29 ages for WHO, 15-34 years for World Bank, and 15-29 years for EU (Global Youth Development Index and Report, 2016). Turkish government policies define the youth within the ages 15 to 24 in Turkey. The official reports state that as of the end of 2021, young population in this age group made up 15.3% of the total population, 51.3% of these people is male and 48.7% is female (Turkish Statistical Institute, 2022). Turkey has the highest percentage of young population among all EU countries, followed by Ireland with 12.6% and Denmark with 12.3% (Eurostat, 2021).

The construction of the *youth* as a category in Western societies can be traced to the emergence of nation-states and industrialization. The period of modernity in Europe in 19th century is closely linked to the manifestation of the youth as a distinct life stage as the rapid industrialization in this period required labour and the labour was provided through longer periods of apprenticeship. Additionally, the license for citizenship was provided via the longer periods of education. As a result, the notion of youth as a distinct category was presented and applied to whom were undergoing the process of apprenticeship and citizenship (Sercombe, 2015). Coined in 20th century, the concept of *teenager* is the product of post-war economic boom in the United States through which young people became the main target and audience of the growing market and popular culture (Neyzi, 2001). The concepts of youth and teenager can be regarded as more of a social and cultural construct. The term *adolescence,* on the other hand, has developmental and psychological underpinnings, it is generally acknowledged that this phase is marked as a physical and biological stage which stars with puberty. (Clark-Kazak, 2009; World Bank, 2007). Nevertheless, it is usually the case that various formal and informal discourses use the terms interchangeably and sometimes simultaneously.

At this point, it is important to mention the generationalist approaches to the conceptualization of youth within the field of sociology in the 20th century. German sociologist Mannheim's influential works (1952) offered categorizations of social generations and the youth was understood in terms of groups of people who inherit ideas from the previous generation and shape the characteristics of their age cohort. Mannheim's theorizations prevailed for a long time, yet it also faced certain criticism over time due to the fact that discursive dimension of generation was largely ignored by Mannheim and this line of thought following him until recently. As Bourdieu (1993) puts it though, youth is a socially and discursively constructed notion which is evident in the struggle between the young and the old. Tendency to think of the whole social order in terms of a scheme of division was a scholastic fallacy for him, he mocked the concept of generations due to the probability that false generalizations can be made based on the attributes of small numbers of elites. Purhonen (2016) also underlines that Bourdieu's insights for generations emphasize that it is a discursive construction.

As for Turkish society, age has always become one of the core cultural constructs in the culture and has been observed in various types of discourses, such as the kinship discourse, nationalist discourses (Neyzi, 2001), and citizenship discourses. In kinship

discourses in Turkey, high value placed upon the concept of seniority is highlighted. For instance in traditional Turkish culture, *elderly* and *juniors* are two distinct categories which is also linguistically reflected on certain kinship terminology (i.e. distinct categories for siblings based on age such as *kız kardeş* for 'sister' and *abla* for 'big sister') and specific honorifics reserved for individuals on the basis of age and gender (i.e., *abi* 'big brother' and *abla* 'big sister' used as address terms also for non-family members). The apparent hierarchy between seniors and juniors and the imposed power and the dominance are legitimized through age and these age-based concepts. In nationalist discourses of Turkey, the construction of youth has developed over time as well. In pre-republic days, the youth was *the hope for the future* to save the country (Young Turks), in the early-republic years they were conceptualized as *the embodiment of the nation* itself, after a period of time they were labelled as *rebels* (68 Generation) and following the 80s, their roles have been redefined as they were *central subjects* within an era of privatization, consumer society and new communication technologies. Today, the fluidity of identities became visible in Turkish public discourses and the youth themselves started to define their own communities and discourses frequently constructed around discourses of citizenship (see section 2.4.1 for the evolving discourses in Turkish youth studies). Never before have Turkish youth challenged the labels assigned to them; however today, they maximize the outreach provided by online tools and manipulate language to challenge the established norms of communication in order to present, project, and negotiate their identities.

In a similar vein, the definitions and the scope of work put forward for language practices of young people have various interpretations. The initial sociolinguistic studies followed the Labovian concept of the *vernacular* to document the routinized and systematic description of regularities in the language of youth. Yet the description of youth language as a distinct vernacular and comparison of it to a mainstream standard language in a society also led to deterministic evaluations of youth language as being deficient, incomplete or transitory language practices. As a vernacular, youth language was seen as a divergence from the base language but it was unmarked and unmonitored. To address the complexity and heterogeneity of language of the youth, Kotsinas (1998) introduced the term *multiethnolect* to depict the linguistic practices of Stockholm youth as a distinct variety along with other varieties in the city. Flourished in Scandinavian sites of research, this variety approach suggested that a multiethnolect is used by the immigrant youth and is characterized by mixing a range of linguistic forms and practices

from various heritage languages with the mainstream language of the local society. Different than vernaculars, though, multiethnolects can be deliberate and marked (Cheshire et al., 2015). In relation to the multiethnolect, Cheshire, Kerswill, Fox, and Torgersen (2011) focused on the linguistic diversity performed by the multiethnic young speech communities in inner-London and coined the term *Multilingual London English* to refer to the repertoire of distinctive cross-linguistic features the speakers make use of. Nevertheless, the term multiethnolect was often criticized for ignoring the performativity of language and implying a positioning with regard to ethnicity. Dorleijin and Nortier (2015), as a response, highlighted the interplay of stylization in the linguistic practices of youth and suggested the term *urban youth speech style.* This issue was also discussed broadly by Rampton (1995) who treated interactional practices of multiethnic youth communities similarly as a stylistic practice and initially defined it as *language crossing* to refer to the ways young speakers of British English cross social and ethnic boundaries by experimenting with various speech styles to manage their relations with peers from different ethnic backgrounds and at the same time challenge inherently ideological stereotypes. More recently, Rampton (2011, 2013, 2015) expanded on the phenomena and adopted the term *contemporary urban vernacular* to encompass diversity of linguistic behaviour such as stylization, crossing, and other meta-pragmatic practices and also to refrain from any references to age.

Even though there is no consensus over the terminology to be adopted, the variety of approaches contribute different perspectives to investigate complementary aspects of the greater whole. It should be noted that recent works advocate for the term *youth languages* to underline the context-dependency, multiplicity, and dynamic characteristics of the discursive interaction particularly among young people. In this view, the common denominator is still the biological age, yet the boundaries of the age spectrum does not have a pre-determined range. In line with this encompassing view, this study also adopts the term *youth language* and -specifically *youth talk* for spoken interaction- to define, explore, analyse its corpus data.

2.2 Research methods in youth language

In this section, two of the most prevalent research orientations adopted in the literature to investigate the youth talk will be presented: variationist studies and corpus driven studies in youth language research.

2.2.1 Variationist studies

The variety approach has its roots in Labovian sociolinguistics. The studies which adopt this approach to youth language investigate the linguistic practices of young speakers as a systematic and structured phenomenon and aim to identify the patterns of variation within their speech. It is argued that variation has a linguistically and socially constrained nature and this view has been investigated in cascading scholarly work which Eckert (2012) categorises as three waves of variationist sociolinguistics. The first-wave paradigm aimed to explain the variation in terms of socio-demographic constraints, particularly social class, gender, and age of the speakers. This line of work which focused on the variation with regard to social class lay ground for the discussions over standard/prestige and non-standard forms of language, which eventually led to the emergence of the view that youth language represents a divergent or deficit form of language (Georgakopoulou & Charalambidou, 2011). In these initial studies, the conception of gender was static and binary (Trudgill, 1974, 1983; Labov, 2001) and the quantitative tradition of this strand of work concluded that women led the linguistic change through their frequent use of new forms in language (Labov, 1966). In terms of the variable of age, the diachronic change in language was explained through the comparison of linguistic patterns across different age cohorts and the studies underlined that the young speakers of language exhibit innovative forms most frequently than other age groups (Labov, 2001, Tagliamonte & D'Arcy, 2009) and the speech of a person gradually becomes more standard in their middle years as they are expected to respond to the speech norms of a particular society (Holmes, 2013). This line of work aimed to achieve cross-linguistic patterns in order to develop a system to predict variation and change across different settings and speakers.

In second-wave studies, the research foci remained the same yet the studies started to make use of naturally occurring data and more qualitatively oriented methodologies such as ethnographies. In contrast to deterministic view of social meaning in the first-wave research, the second-wave studies highlighted the speaker agency in vernacular use. Labov's (1972) study, for instance, showed that vernacular use of the young speakers of African American Vernacular English in New York indexed their in-group status. Cheshire (1982) conducted a study on vernacular use in a working-class youth sub-culture and illustrated that non-standard morphosyntactic patterns in their language implied different social norms. Eckert's (1989, 2000) ethnographic studies are

also among the representative work on phonological variation observed in the talk of two distinct groups of Detroit youth called 'Jocks' who consist of middle-class youth and 'Burnouts' who belonged to working-class. The results of these studies indicated that each phonological variable in the study correlated with gender or the social class. Additionally, social category affiliation intertwined with different social norms and beliefs provided explanations for the phonological variation observed across these groups.

While second-wave research explored the influence of context and social categories over linguistic practices, it is the third wave of variationist studies which particularly focused on the dynamic and tailored stylistic practices of speakers. This line of studies specifically explored the issues of identity and ideology (Eckert, 2008) and pointed out that identity is a dynamic and fluid concept and ideology can be reconstructed and reproduced through discourse (Agha, 2007; Svendsen, 2015). The research within third-wave sociolinguistics is more interested in the social meanings, functions, and consequences of the youth talk (Quist, 2008). In this line, the recent scholarly work in third-wave variationist paradigm mainly explores young people's stylistic preferences such as the use of slang or so-called non-standard uses of language not as a deficit vernacular use –as implied through first-wave studies- but as a way to show that young speakers can consciously manipulate the language and the speech styles in accordance with the relevant context and interactional goals (Androutsopoulos, 2015; Bodén, 2004, 2011; Eckert, 2000; Ilbury, 2019; Jørgensen, 2008; Madsen, 2015; Sierra, 2016).

2.2.2 Corpus driven studies

In addition to the variationist approach which traditionally makes use of more ethnography-oriented methodologies to study youth language, there have been a growing body of literature which use corpora both as a methodology and as a source of data to investigate youth language for the last 30 years. While sociolinguistic tradition and corpus linguistics are not mutually exclusive approaches to the study of youth language, it is important to present the projects which were designed, compiled and constructed with the specified purpose of examining the linguistic practices of young speakers of various speech communities. While these projects have overlapping research agendas with variationist studies, they stand out within the youth language research in terms of their sustainability (e.g., use of concordancing and monitor corpora),

cross-linguistic comparability, and offering a representative sample of the community they present. To elaborate on the scope of work conducted at the intersection of corpus linguistics and youth language studies, the prominent spoken youth corpora built for English, Spanish, German, Danish, Finnish, Icelandic, Norwegian, and Swedish will be presented in the following section. Additionally, major learner corpora with pedagogically driven goals to study youth talk will also be exemplified.

2.2.2.1 English youth talk corpora

Over the past thirty years, there have been growing increase in the use of corpus methods utilized for the study of youth language. Currently, there are corpora of various sizes focusing on youth talk in various languages. The pioneering work was carried out by Stenström and her team (Stenström et al., 2002) who built the first spoken youth language corpus, namely The Bergen Corpus of London Teenage Language (COLT). The researchers collected audio data from 33 English speaking teenagers between the ages (in majority) 13 to 17 of various London boroughs in 1993. The participants were coded for age, gender, social class, ethnicity, setting and location. The COLT Project was an international collaboration and was supported by several funding bodies. The project was carried out by researchers at Bergen University and received assistance for transcription from Longman Group, word-class tagging by Lancaster University, and technical support from Norwegian Computing Centre for Humanities. The data collection followed the design of the BNC while the sampling was restricted solely to the London area rather than the whole of Britain. A total of five London school boroughs were chosen on the basis that each represented one social class. Schools helped the researcher team to find recruits to make recordings. This 444,166-word corpus was later incorporated into the BNC1994 and currently is available for academic purposes upon request[7]. The construction of the COLT enabled researchers to conduct linguistic analyses on a wide range of linguistic devices (e.g., discourse markers, swear words, slang expressions, intensifiers, tags) and phenomena (e.g., mimicry, conflict talk, storytelling) observed in English language spoken by young speakers. In their comprehensive work on the COLT data; Stenström, Andersen and Hasund (2002) note the use of 'slanguage' which consists of slang words, swear words, vogue words, vague words, set markers, quotatives, hedges, empathizers, and tags as the most salient feature of British teenage talk.

[7] Please visit http://clu.uni.no/icame/colt/ to access the corpus.

Additionally, playful, creative and innovative use of language is highlighted as a manifestation of identity expression for young Londoners. Following this preliminary yet elaborated investigation of the COLT data, several researchers have made use of the corpus in their own separate analyses (Andersen, 1997, 1998; Drande, Hasund, & Stenström, 2014; Drummond, 2020; Hasund & Stenström, 1997; Palacios Martínez, 2011a, 2011b, 2018; Rodríguez González & Stenström, 2011; Stenström, 1997, 1998).

Looking back at the reports regarding corpus compilation stage of the COLT project (Stenström et al. 1998), the corpus bears a few limitations in terms of its sample. Though it is a corpus of youth talk, the data does not exclusively have teenager talk but rather it includes dialogues between teenagers and young adults or adults. The group of young adults (between ages 20 to 29) make up 0.28% of the COLT while the group of adults above age of 30 include parents and teachers correspond to 6% of the corpus data overall. Secondly, it is also reported that the metadata regarding socio-economic background of the participants were merely coded for 50 percent of the speakers.

The Toronto Teen Corpus (TTC) consists of 90 speakers between the ages 9 to 22 from different education levels in Canada (Tagliamonte, 2016a). The spoken data was collected between 2002 to 2006 and is reported to be nearly 1 million words. It is reported that the spoken data was collected through interviews conducted by a group of undergraduate researchers. The researchers interviewed their friends, siblings, cousins, and neighbours and the topics were mainly guided by the researchers. There is no information regarding the scope of metadata or the annotation scheme of the corpus. Tagliamonte (2016a) reports that she also collected successive corpora called The Toronto Instant Messaging Corpus (TIMC) and The Toronto Internet Corpus (TIC) for which participants also consisted of Toronto youth. The main goal of these projects was to identify the innovative changes and variation in Canadian English and utilizing corpus tools, Tagliamonte (2005, 2008, 2016a, 2016b) explored a wide range of linguistic devices and phenomena such as quotatives, intensifiers, discourse markers, adverbials and adjectives, and general extenders. Drawing from a range of data sources and methodological tools, Tagliamonte's corpus investigations revealed both evidence for language variation across age groups (e.g., the results showed that young speakers use quotatives such as *like, say*, and *go* for different pragmatic purposes than adults), within-group (e.g., young female speakers are reported to use intensifiers more frequently than males), and across time (e.g., *like* and *so* are observed more frequently in youth data

while *you know* is more used by those born in 1975-89). In this respect, TTC is a comprehensive and rich resource for the studies of youth talk in Canadian English.

2.2.2.2 Spanish youth talk corpora

Regardless of its limitations, the COLT paved the way for the successive corpus projects focusing on youth talk. Among them is Corpus Oral de Lenguaje Adolescente (COLA) which was built to explore Spanish youth talk. The project was led by Annette Myre Jørgensen and Anna-Brita Stenström from the COLT project, and was funded by University of Bergen, Meltzer Fund, and Research Council of Norway. The corpus was compiled from 145 young speakers of Spanish between ages 13 to 18 from Spain. The corpus has three sub-corpora: COLAm consists of youth talk from Madrid, COLAba is the corpus of youth talk in Buenos Aires, and COLAs includes youth talk from Santiago de Chile. The corpus data covers the period of 2002-2004 and 2007, and makes up of 500,000 words in total. The setting of the recordings included both school and places outside of the school such as homes or parks. The speakers were coded for age, gender, social class, and type of school. Using COLA, linguistic features of Spanish speaking youth such as discourse markers, anglicisms, intensifications and taboo words have been investigated (Drange, 2009; Stenström, 2007, 2014; Stenström & Jørgensen, 2009; Jørgensen, 2008, 2009, 2013).

COLA is prominent in the sense that it not only provided corpus tools for the systematic analysis of a language other than English but also enabled cross-linguistic analyses in youth talk research. The COLT and the COLA projects both followed the same pattern of data collection and corpus structure used for the BNC, thus it is possible to do cross-linguistic research between English and Spanish. For instance, Stenström (2005, 2014) compared the youth talk in English and Spanish focusing on the use of taboo words, pragmatic markers such as address terms, intimacy markers, intensifiers, hedges, slang, and also politeness. It is revealed that Madrid girls between the ages 14-15 with a middle class background are the most frequent users of pragmatic markers while in the COLT, boys between the ages 14-19 with a high class background use pragmatic markers more often than girls. Additionally, null/zero quotatives, namely the absence of introducing verbs in direct speech, is identified as an important element in the construction of both Spanish and English teenagers' narratives in the COLT and the COLA (Martínez, 2013).

Recently, the Corpus Oral de Madrid (CORMA) corpus was built to document linguistic characteristics of contemporary spoken Spanish (Enghels et al., 2020; Roels, 2021). The CORMA corpus was compiled by researchers at Ghent University department of Spanish Linguistics. It is a 476,606-word spoken corpus which contains 529 speakers from four age cohorts (grouped into 0-11, 12-25, 26-55 and above 55) and socioeconomic backgrounds (grouped into high, middle and low) in Madrid, Spain. The age cohort of 12-25 year-olds is labelled as 'GEN2' and this particular group corresponds to the sub-corpus for adolescence speech which makes up 154,117-words from 139 speakers. The corpus contains five distinct communicative settings of family, friends, customer service, colleagues and acquaintances. Utilizing the affordances of CORMA, Roels and Enghels (2020) investigated age-based variation in intensifying strategies in Spanish across age groups and demonstrated that speakers from younger generations incorporated higher number of intensifiers to their speech. In terms of the types of intensifiers used, younger speakers made use of more expressive types of intensifiers which is argued by the study as the reflection of linguistic innovation. Thanks to distinct age cohorts in the design of the corpus, CORMA offers a sub-corpus of spoken Spanish which is comparable to COLAm. This comparability enables monitoring the linguistic change in Spanish youth talk across time. In this line, Roels, De Latte and Enghels (2021) focused on use of vocatives and intensifiers within a period of fifteen years, and demonstrated that linguistic changes do occur over time yet in a moderate speed. Based on further investigation, the researchers correlated speed of change with two principles. It is proposed that standardized forms tend to remain stable over time and that more expressive types are picked up and abandoned quickly. These results are valuable in the sense that they contribute robust evidence to the previous arguments that younger speakers lead the linguistic change and the results lays ground for the forthcoming diachronic corpora studies.

As new modes of communication have become a vital part of daily youth interaction, some researchers adopted corpus tools to develop corpora of language used in online spaces. An example is Dorantes, Sierra, Perez, Bel-Enguix, and Rosales' (2018) the Sociolinguistic Corpus of WhatsApp Chats which consists of 756,066-token written and spoken data from Spanish speaking college students in Mexico City. It is reported that of the total of 1325 informants, 84.9% is undergraduate students. It is important to note that the corpus data is not exclusively youth language as the researchers did not exclude conversations between college students and their families or co-workers. As a result, the

age of participants ranges between 14 to 60. The project intends to investigate the virtual interaction via Instant Messaging (IM) among undergraduate students via the most frequent lexical words, emoticons, parenthetical expressions, code-switching, turn-taking, speech acts, linguistic variations identified in the corpus.

2.2.2.3 German youth talk corpora

As for German youth talk, there are three distinct spoken corpora built with different research foci. The Ph@ttSessionz Project (Draxler et al., 2008) aimed to build a 'database' of read and spontaneous speech from 864 speakers of German between the ages 12 to 20. The database is a sub-corpus of The Regional Variants of German Corpus (RVG-1) constructed by researchers at University of Munich (Burger and Schiel, 1998) and was funded by The Federal Ministry of Education and Research in Germany. The data was collected between 2005 to 2007 through online data collection tools which allowed researchers to reach out to a high number of participants from a variety of public secondary schools across Germany. The participants were either required to read the prompts or provide unscripted answers to the prompts. The project reports the size of the corpus in terms of 'utterances' recorded, it is noted that it consists of more than 110,000 utterances. As the project is interested in influence of age and gender over phonological features of speech, the collected metadata focused on demographic information regarding the dialect region of speaker, mother tongue of speaker and their parents, as well as the details about oral health (e.g., smoking habits, lip and tongue piercings, braces) of speakers. The corpus stands out as a specialized spoken corpus built for serving speech recognition technology and spoken dialogue systems. The Voices of Young Scots (VOYS) is a project with the same data collection design and sociophonological research agenda. With the cooperation of research team of The Ph@ttSessionz Project, the VOYS was compiled in 10 locations across Scotland and it consisted of 300 young speakers of Scottish between the ages 13 to 18 (Dickie et al., 2009).

The KiezDeutsch-Korpus (KiDKo) was developed at the University of Potsdam (Rehbein et al., 2014) from 2008 to 2015 and was funded by German Research Association. KidKo is a multimodal corpus which consists of five sub-corpora. KiDKo/Mu and KiDKo/Mo correspond to the main corpora of 333,000-word spoken corpus which includes spontaneous spoken data of informal conversations among friends who are between 14

to 17 ages from two boroughs of Germany. The language spoken is mostly German, along with codeswitching instances of Turkish, Arabic, and Kurdish. The total number of speakers in the corpus is 23 and the scope of metadata recorded for each speaker include gender, residential area, and family language. The focus of this corpus is the language practices of young people among their peers in multiethnic (KiDKo/Mu) and monoethnic (KiDKo/Mo) residential areas, document linguistic developments in contemporary German, and explore youth language as an informal urban use of language. The other sub-corpora are KiDKo/LL which is a corpus of photos of written data captured on walls, park benches, graffities in urban and public spaces and KiDKo/E which is a corpus of emails and letters regarding language attitudes and ideologies. There are also three smaller supplementary corpora which have spoken data elicited through storytelling or language situation prompts.

Jugendsprache Schweiz Korpus (JuBE) was constructed by a research team at the Center for the Study of Language and Society at University of Bern, Switzerland. The corpus data was collected from 26 German-speaking youth between the ages 12 to 22 in the canton of Bern between the years 2019 to 2020, and the project is currently at the stage of data transcription. Similar to the KiDKo project, the main goal of the JuBe project is to investigate the linguistic innovation, change, multilingual practices in youth talk as well. The project puts emphasis on exploring the 'ethnolects' in youth speech as Switzerland has four official languages and languages of immigrants are also salient in the society (Schneider et al., 2021).

2.2.2.4 Nordic youth talk corpora

The increasing exposure to different languages and cultures has led researchers in the Nordic countries to adopt corpus tools to study the language change and variation observed through youth language, as well. The most comprehensive research on youth language was carried out by the UNO Project - Språkkontakt och ungdomsspråk i Norden (Nordic Teenage Language) which collected data from students between 13 and 19 from Denmark (446 students), Finland (481 students), Iceland (1226 students), Norway (422 students) and Sweden (2105 students) between the years 1997-1998. UNO relied on data from a comprehensive written slang survey and self-recorded conversations. The main foci of the project were the slang expressions and discourse markers. In terms of youth slang, Drange (2002) traced the origin of expressions in Norwegian youth slang

and reported that 20 different languages were identified for these borrowings. Among these languages, 20% of the slang expressions out of 22,000 words was from English and 3% was from Arabic and Spanish. It is indicated that many of the borrowings undergo the process of adjusting to Norwegian spelling and morphology. The study also presents a brief comparison between Swedish and Norwegian slang and points out that the most frequently used types of slang expressions are different for these languages. Within this overarching project, UNO-Oslo corpus is a 206,854-word spoken youth talk corpus yet it contains some data from adult speakers just like the COLT and COLA. It consists of 18 hours of self-recorded conversations from 45 participants (Hasund & Drange, 2014). The speakers in UNO-Oslo were coded only for two socioeconomic backgrounds, namely middle-class and working-class. In a complementary study, Drange, Hasund and Stenström (2014) compared swearing practices observed in English, Spanish and Norwegian through the COLT, COLAm and UNO-Oslo corpora. Even though the corpora are not completely comparable in terms of several levels such as size, time period, sociolinguistic distribution of speakers; the study illustrates the affordances of corpora to track the discursive associations of swearing practices in different languages.

As a separate initiative, there is also The Icelandic Spoken Language Corpus (ISLC) which contains four different sub-corpora of spontaneous conversations, group conversations, parliamentary debates and conversations of teenagers. The sub-corpus of teenager conversations was collected through the project *How do young Icelanders speak in the beginning of the 21st century?* in 2006 and was funded by University of Iceland and the Ministry of Education and Culture. The corpus consists of 35,527-tokens in total and was tagged morpho-syntactically. It is reported, though, that the interactional data in the corpus is not limited to peer talk among youth but rather there are adult speakers in conversations as well. The main goal of the project is to build a linguistic resource for Icelandic language technology projects (Steingrímsson et al., 2018).

2.2.2.5 Spoken learner corpora

In addition to the specialized corpora of aforementioned languages which were built as representative projects with the purpose of examining the linguistic practices of young speakers within informal domains of interaction, it is also worth mentioning the youth language corpora which were exclusively compiled from the context of language learning.

Among the major projects include International Corpus of Learner English (ICLE) which is currently a 5.5 million word corpus from learners of English across 25 mother tongue backgrounds. The project was initiated by The University of Louvain and the data was compiled through collaborations with partner universities across the globe. The written corpus comprises 9,529 essays of students of English (Granger et al., 2020). The corpus includes the metadata of age, gender, mother tongue background, region, knowledge of other foreign languages, time spent in an English-speaking country, learning context, and proficiency level. The age of the participants ranges from 16 to 71, with an average age of 22.27 in the whole corpus. The corpus has played a key role in promoting the affordances of building and using learner corpora. Following ICLE, The University of Louvain widened its agenda of corpus projects and built several other types of learner corpora. Among them, Louvain International Database of Spoken English (LINDSEI) is the spoken counterpart of ICLE and contains spoken data from advanced learners of English who are undergraduate students with different mother tongue backgrounds. Constructed using the same design with LINDSEI, New Englishes Student Interviews (NESSI) corpus has informal interview data from young speakers of New Englishes and aims to offer comparisons between New Englishes and learner Englishes.

The System Aided Compilation and Open Distribution of European Youth Language (SACODEYL) is an EU project which aims to construct spoken language corpora from English, French, German, Italian, Lithuanian, Romanian and Spanish youth talk. Similar to LINDSEI and NESSI, the main goal is to offer a pedagogical resource for language learning/teaching and facilitate data-driven approaches to language acquisition (Pérez-Paredes & Alcaraz-Calero, 2009). In SACODEYL, the speakers are between the ages 13 to 18 and the corpus data was compiled though interviews which took 10 minutes for each participant. The data was elicited through pre-determined set of topics and questions posed at participants who talked either individually or in pairs, and it is reported that the corpus has 20 to 25 video-recorded interviews. Though the common denominator for these corpora and the youth talk corpora is age, it should be noted that these corpora are not designed to be representative of youth talk but rather 'be pedagogically representative of the type of language required by teenage language learners' within their context (Pérez-Paredes, 2019).

Other notable spoken learner corpora include The Corpus of Young Learner Interlanguage (CYLIL), The College Learners' Spoken English Corpus (COLSEC), The Tübingen Corpus of Eastern European English (TCEEE), Evaluation of English in Norwegian Schools (EVA), and The Finnish Upper Secondary School Spoken English (FUSE).

The Corpus of Young Learner Interlanguage (CYLIL) contains longitudinal spoken data of 500,000 words from English learning European school pupils from Dutch, French, Greek, or Italian mother tongue backgrounds. The project elicited data through both interviews and informal semi-structured conversations with the purpose of documenting language learning development process (Housen, 2002). The College Learners' Spoken English Corpus (COLSEC) which is a 700,000-word corpus constructed to generate pronunciation lexicons and avoid mispronunciations (Yang & Wei, 2005). The Tübingen Corpus of Eastern European English (TCEEE) which has 60,000-words of spontaneous spoken data from Slavic speakers of English with Ukranian, Russian, Polish or Slovak mother tongues and constructed in order to investigate the morphosyntactic and morphosemantic features of the expanding circle Englishes (Salakhian, 2012). Evaluation of English in Norwegian Schools (EVA) is a 35,000-word spoken corpus which consists of 14-15 year-old Norwegian pupils' oral test transcriptions. The EVA was designed similar to the COLT corpus to allow comparisons (Hasselgren, 2000). Additionally, there is The Finnish Upper Secondary School Spoken English (FUSE) corpus which consisted of spoken conversations recordings of students who took Oral Examination in English. Following the same design of the SCOTS corpus, the data collection for the FUSE project started in 2014 and aims to grow in size over time (Ehrnrooth, 2015). It is reported that the corpus currently has 20,329 words in total (Lukkari, 2020). The main goal of the corpus is to provide a linguistic resource for teachers and learners of English.

2.3 Recent foci of investigation

In line with the research orientations and methods presented, the body of literature on youth talk cluster around a number of research themes or foci of investigation. In this section, a selected number of recent research foci will be presented. These include studies focusing on the characteristics of youth talk within the scope of (i) indexing

identities, (ii) linguistic innovation and change, (iii) multilingual encounters, and (iv) stylization in digital sphere.

2.3.1 Indexing identities

Linguistic studies on youth language highlight that one of its prevailing functions is fostering in-groupness. Jørgensen (2013) indicates that teenagers' use of particular talk is a way of articulating their identity independent of adults and creating a bond with their peers. It is indicated that peer-group identity is considered important for teenagers, thus the use of a particular speech type fosters both the individual and the in-group identity. It is possible to observe in-group variation within the same group of teenagers in a community as well. Madsen (2013) highlights that social power differences can be tracked in the linguistic features of Danish youth talk. Speakers' ideologies in relation to the concepts of 'insider' and 'outsider' are reflected in the distribution and change of linguistic forms of urban speech of Copenhagen youth, thus an ethnographic perspective to youth talk is advocated. In a similar vein, drawing upon Foucault's (1977, 1980) concepts of power and knowledge in interaction, Irwin (2006) discusses the co-construction of identity in working-class versus middle-class London youth based on spontaneous speech data. It is argued that the pragmatic expressions *you know* and *I know* signal the potential social positioning of the self. The study illustrates that *you know* is mostly used by working-class London teenagers and represents a relatively active identity construction while *I know* is used mostly by middle-class London teenagers and relatively reactive identity construction.

Drummond's (2016) The UrBEn-ID (Urban British English and Identity) project is an ethnographic work on young people's language practices and identity enactments in Manchester, UK. The data was collected in 2014-2015 and 70 hours of audio recordings along with 413,000 words of fieldnotes were compiled. Audio recordings consisted of spontaneous conversations between 14-16 year old speakers of English as well as conversations and interviews between the participants and the researcher. One of the contributions of this project is to illustrate the way young people view their language practices. The study showed the youth language incorporates various linguistic and semiotic resources and young speakers of English are capable of adjusting their language use based on the contexts, and thus different identities are dynamically constructed in interaction. In a complementary study, Drummond (2018) demonstrated that the

specific linguistic feature of th-stopping which is generally associated with black varieties of English is not a marker of ethnicity in urban youth language, but rather it is an in-groupness marker indexing a specific youth sub-culture in Manchester.

The constant reconstruction of identity is also discussed by Harissi, Otsuji and Pennycook's (2012) work on spoken interaction data between Greek youth. The researchers indicate that the interplay of different cultural and linguistic repertoires utilized in discourse can be investigated in relation to performativity. The results show that young speakers engage in fluid cultural identifications which is shaped by the discourse. In another ethnographic study on 14-15 year old male speakers in Glasgow conducted between years 2005 to 2007, Lawson (2011) identified patterns of variation in linguistic practices across three distinct communities of practice (CofP) in the data. Speakers aligned their speech in accordance with their CofP membership and position themselves distinct from the out-groups. In Moore's (2004, 2006) studies, which also adopted CofP framework, young female speakers of English from different social groups adopted divergent linguistic practices. The study consisted of spoken data and fieldnotes collected from 40 high school students in England, and the participants aged 12-13 years when data collection started in 2000. The results show that youth manipulated the use of nonstandard grammar and tag questions to create social meanings in relation to emphasizing their positions within their in-groups.

While there are studies which investigate the general characteristics of youth talk with regard to indexing identity and showing affiliation to the in-group, some researchers prefer to focus on specific linguistic devices such as swear words (Palacios Martínez, 2011a, Stenström & Drange 2014) vague language and intensifications (Palacios Martínez, 2011b, 2014; Palacios Martínez & Núñez Pertejo 2012, 2015), vocatives (Palacios Martínez, 2018; Rendle-Short, 2008) to explore this issue. These studies investigate the formal characteristics as well as patterns of these linguistic devices and highlight the function of fostering in-groupness as a salient socio-pragmatic function in interaction.

2.3.2 Linguistic innovation and change

Variationist research focusing on youth language naturally highlights the aspects of language innovation and change in their studies. Since the seminal work of Labov (1992)

which underlined that the variation within youth talk should be analysed in order to track the maintenance, diffusion, or extinction of specific slang terms over time; several studies have demonstrated that the language of younger speakers diverged from the speech of other age groups. It is indicated that the speech of a person gradually becomes more standard in her/his middle years as s/he has to respond to the speech norms of a particular society (Holmes, 2013). Among the studies exploring the divergences young speakers exhibit in their linguistic practices, Palacios Martínez (2011a) compared several features of teenagers' language from the COLT and the SCOSE (The Saarbrücken Corpus of Spoken English) with that of the language of adults from the DCPSE (The Diachronic Corpus of Present-Day Spoken English) along with supplementary data from teenager magazines, web-based glossaries and dictionaries. The study explored a wide range of lexico-grammatical elements and identified the features which distinguished youth talk from the language of adults speaking British English. These features included frequent use of swear words as vocatives, the use of quotative *go* and *like* in reported speech, using placeholders, approximators and general extenders as frequent forms utilized for vague language, using adjectival and adverb intensifiers, and the use of non-canonical tags such as *right, innit*; and vernacular negative forms such as *ain't, nope, dunno.*

Echoing Eckert's (1997, p. 152) famous statement "Adolescents are the linguistic movers and shakers […] and as such, a prime source of information about linguistic change and the role of language in social practice", researchers vastly explored and identified the recent developments and innovations in the language observed through the lens of youth talk. These studies indicate that new forms or pronunciations to express a particular concept evolve over time and youth talk is a resource for identifying the emergence of these new forms or uses in a language (Cheshire, Kerswill & Williams, 2005; Torgersen, Kerswill & Fox, 2009; Holmes, 2013). Studies on youth talk also explored the use of new forms and interactional strategies adopted by young speakers as a result of language contact. These studies explore the issues of multilingualism, multiethnolects, stylization, as well as the concepts of ethnicity and gender. The following section provides a brief overview on this particular strand of research.

2.3.3 Multilingual encounters

Investigating youth talk with relation to multilingual encounters initially flourished in multilingual and/or multiethnic contexts where fluxes of immigration started to shape the linguistic practices in various parts of the world, specifically in Europe. These initial studies either focused on the features of language of immigrant youth such as the language of German speaking Turkish descent teenagers in Germany (Keim, 2001), or the influence of immigrant languages over the linguistic practices of local youth such as the use of Turkish by young speakers from German and other ethnic backgrounds in Germany (Auer & Dirim, 2001). In both parties of work, fluid identities constructed as a response or resistance to stereotypes, ethnicity, and hegemony are explored. In Jonsson's (2018) work on linguistic styles of Swedish speaking male immigrant youth, harmonious interaction among young speakers is fostered though the incorporation of urban youth styles into teaching and the speakers make use of linguistic practices associated with otherness as a resource for conversational humour. In Rampton, Charalambous, and Charalambous' study (2014), on the other hand, it is illustrated that multilingual encounters are not always welcome. The study explores the strategies language teachers adopt in order to refrain from inducing hostility while teaching Turkish in Greek-Cypriot context. It is indicated that Turkish learning and speaking 16-17 year old Greek-Cypriots were negatively labelled in their local context due to the legacy of post-conflict era.

Studies on multilingual youth practices is specifically prevalent in Nordic countries due to the increasing number of multiethnic and multilingual communities in recent years. In Pharao, Maegaard, Møller, and Kristiansen's (2014) work in multiethnic settings in Copenhagen, the researchers showed that depending on the registers, young speakers of Danish associate different ideological schemes with the same phonetic feature. Quist's studies (2008, 2010) approach the language use and variation in the bilingual youth of Copenhagen from a multitude of sociolinguistic perspectives such as stylistic practice approach and variety approach. The results illustrate that there is no direct relationship between ethnic background of young speakers of Danish from immigrant backgrounds and their use of multiethnolects, and that the multiethnolect is in constant interaction with the broader linguistic landscape. As a result, it is suggested that plurality in approaches is a precondition to study ethnolects in the Scandinavian contexts. There are also studies which explore trans-Scandinavian multiethnolectal patterns (Quist &

Svendsen, 2010; Svendsen & Røyneland, 2008) among youth talk in multilingual urban settings.

Rampton's (1995, 1998, 2006) line of work has been prominent since he introduced the term *language crossing* in his study on young speakers in multicultural and multilinguistic urban spaces in British secondary schools. His studies showed that young speakers performed language practices which are "not generally thought to belong to the speaker" (Rampton, 1998, p. 1). Rampton drew from interactional sociolinguistics and ethnography to show that young speakers of English tailored their speech styles with the purpose of maintaining relations with their peers from different ethnic and linguistic backgrounds. Later Rampton (2011, 2013, 2015) suggested the term *contemporary urban vernacular* without restricting these linguistic practices to young people. Rampton's works lay ground for the growing body of research on stylization in digital sphere as will be presented in the following section.

2.3.4 Stylization in digital sphere

The more recent studies, on the other hand, increasingly approach the youth styles from the perspective of new communication tools observed in the digital sphere (Androutsopoulos, 2007; Dovchin, Pennycook & Sultana, 2018; Georgakopoulou, 2008, 2016, 2019; Illbury, 2022a, 2022b; Nortier, 2016; Nørreby and Møller, 2015). These studies highlight that the youth heavily engage in and manipulate multimodal linguistic and semiotic resources while they interact with their peers. Studies have shown that youth appropriate their language in digital sphere which leads sociolinguistic style transfer into a digital style. Androutsopoulos (2007), for instance, showed that the discourse of German speaking youth showed variation in terms of vocabulary, discourse markers and spelling to negotiate their online identities which can be a reflection of their offline identity or a totally alternate online one. Nortier (2016) investigates the variation from a perspective of stylization and contributes to Androutsopoulos' (2007) work by revealing that youth can adopt different linguistic styles in different genres of online communication channels.

In this line, Nortier (2016) explores multiethnic urban youth style in Netherlands by investigating a rap video of a group of young Moroccan- and Turkish-Dutch rappers and discusses the variation in stylization observed though the use of different accents by the

speakers in the video as opposed to their interview recordings. Studies also explore the effect of online discourses over offline interactions among youth. Rørbeck Nørreby and Spindler Møller (2015), for example, shows how online recourses presented in social media trends influence the societal discourses formed around the concepts of beauty and desirability with relation to ethnicity influence over the everyday interactions of urban youth in Copenhagen. More recently, Illbury's works (2020, 2022a, 2022b) explore the linguistic patterns in the offline-online interface and discuss the reciprocal interaction between digital culture and everyday language of young speakers of English. Illbury (2022a) argues that youth view constructed discourses in social media as an extension of their offline social network in his study where he investigated the trends and discourses regarding different types of social media used by East London youth. In another study focusing on youth language on Instagram, Illbury (2022b) explores stylization in digital contexts and demonstrates that speakers construct stylistically adjusted digital identities via feeding from semiotic resources such as memes which both reflect specific language ideologies existing in offline discourse and generates new indexical and ideological associations.

It is also important to note Georgakopoulou's works on youth talk (2008, 2016, 2019) which draw from small stories analysis (Georgakopoulou, 2007) as an alternate narrative analysis approach to investigate the situatedness of interaction. In her ethnographic work on 14-15 year old students in London (Georgakopoulou, 2008), for instance, self- and other-identity claims of the youth is investigated in mediated interaction through MSN texting data. The study suggests that rather than 'big' classifications of identities such as ethnicity, speakers' focal concerns should be explored. In a successive study, discourse of *selfies* of 16-18 year olds were examined (Georgakopoulou, 2016). The results show that youth perform interactional alignment patterns with regard to the created stories via selfies. As a complementary finding, Georgakopoulou's (2019) corpus assisted work underlines the affordances of tools such as *stories* presented by specific social media platforms which presents young adults new forms of self-presentation.

Additionally, Dovchin, Pennycook, and Sultana (2018) draw from the concepts of translingualism and heteroglossia and investigate the multilingual practices of youth from Asian peripheries in both online and offline context. The study offers a 'transglossic'

framework to analyse the ways young speakers utilize the semiotic resources feeding from global popular culture and construct relations with the issues in their local context.

2.4 Youth studies in Turkey

Though Turkish youth studies is not a new research area, the scholarly research within the field of language studies is fairly limited. The bulk of research on youth was carried out within the fields of anthropology, psychology, sociology, education, history, sport sciences and political sciences. In this section, firstly the guiding studies from the informing fields of education, psychology and sociology will be used to outline the historical development of youth studies in Turkey. Then, the studies conducted in linguistics and language teaching will be presented.

2.4.1 Informing fields

In his comprehensive systematic review of scholarly articles, masters and doctoral theses, and books published between 1923 to 2012 in Turkish academia, Yaman (2010, 2013) offers an account of salient themes and trends in youth studies in Turkey. It is reported that the majority of youth studies work belongs to the fields of education, followed by psychology and sociology in which the label *youth* referred to the university students for the most part. The review shows that the studies within the field of education densely clustered within the time period 1923 to 1950 which corresponds to the foundation and the early years of the Turkish Republic when the state prioritized transmitting the national goals and ideologies to the youth. Kaplan (1999) indicates that within this period, education was a key to create a homogenous society with a monolithic identity. In this line, the youth was central to this agenda as the newly established state needed "a new type of person with a new mind-set" (Neyzi, 2001, p. 416). As a result, the studies published within this period did not regard the youth as the object of scholarly investigation but merely as the target audience. Rather than conducting ethnographies or administering any kind of surveys, studies in this period took a prescriptive approach and treated the youth as a container to tuck in a pre-determined collection of ideal values, attitudes, behaviours, and vision in line with the discourses of nationalism. The educated youth was regarded as the representative of the Turkish nation as a whole and the youth was used as a political agent to ensure that the nation would catch up with the Western scientific knowledge and defend the national virtues at the same time (Yolcu,

2014, 2019). The intellectual and moral transformation of the youth were the primary goal of the studies published (Demir, 2012).

In the second phase of youth studies which spans the time between 1950 to 1980, the scholarly literature progressed within the fields of psychology and sociology (Yaman, 2010) in line with the *sociocultural* (e.g., migration of rural to urban spaces, labour-migration to Europe), *economic* (e.g., rapid industrialization), and especially *political* developments (adopting a multi-party system, university student movements) experienced in Turkey. As a result, the scope of youth studies diversified and focused on issues such as politicized views of university students (e.g., Abadan Unat, 1961; Ozankaya, 1966), structure of youth movements (e.g Bulutay, 1969; Kışlalı, 1972, 1974), and urbanization and its effects on youth development (e.g., Gökçe, 1976; Yörükoğlu, 1968). This period also marks the start of the systematic research on Turkish youth as the researchers started to adopt various data sources and methodologies such as field works, qualitative and quantitative analyses into their investigations (Yaman, 2013).

From 1980s onwards; the 1980 military coup, privatization, the rise of the consumer society, new communication tools and developments in technology led the youth studies in Turkey to take a discursive turn. While the youth was regularly defined as active political agents or potential threats in the 1970s in Turkey (Doğanay, 2018), opinions about their representations diversified in the post 1980s era. Various alternate labels were associated with the youth such as individualistic, liberal, apolitic, apathetic (İMV-SAM, 1995). In a comprehensive study administered to 2223 young people between the ages 15 to 27 from 11 provinces in Turkey, Konrad Adenauer Foundation (1999) explored the attitudes and values of youth regarding local and global sociopolitical issues, positionings of self and the other, as well as the current problems, concerns, visions of the young people at that time. The results of the survey state that the majority of the Turkish youth viewed their generation as "aimless, idealless, insensitive" (Konrad Adenauer Foundation, 1999, p. 47). Around this time, the studies started to widen the spectrum of the scope of the sample for the youth and included young people in high schools among their target population of inquiry, as well. For instance in 1981, Tezcan's work put spotlight on high schoolers as another alternate representative sample for youth population in Turkey. A closer look into the masters and doctoral theses published between 1980 to 1990s which had high school students as their sample shows that studies within the fields of education and psychology mainly focused on academic

achievement of students (e.g., Erdoğan Baydilli, 1987; Şengil, 1985), test anxiety (e.g., Çankaya, 1997; Sargın, 1990) and English language learning difficulties (e.g., Akufuk, 1996; Kafes, 1998; Köprülü, 1994). Works on sociology, in the meantime, explored the attitudes of youth regarding social, cultural and political activities (e.g., Kentel, 1995; C. Kozanoğlu, 1992; H. Kozanoğlu, 1993), child labour, street children, alcohol and drug addiction (e.g., Küntay, 1999; Tütün, 1995) and youth sub-cultures (e.g., Burcu, 1997; Doğan, 1994). It is also important to note that by the late 1980s, diaspora youth became one of the research topics in social sciences in Turkey (Demir, 2021). Looking back at the methodologies adopted to investigate these topics, Demir (2012) notes that studies conducted by critical sociologists in Turkey in 1990s were carried out within the frameworks of critical theory and cultural studies of 1970s and the boundaries of European youth research, thus the analytical approaches and theoretical concepts failed to provide Turkish researchers with in-depth analysis tools attuned to their local context. It should be noted that Kağıtçıbaşı (1984, 1996) contributed immensely to the study of self and the assumptions about the socialization of youth through her studies on convergence hypothesis within the field of cultural psychology. Her studies challenged the Western assumptions regarding self and development and she revealed that autonomous goals of the youth does not necessarily imply emotional interdependence of youth from their parents.

Demir (2012, p. 98) indicates that youth studies between 1980-2000 in Turkey mainly ignored contextual variations and interdisciplinary models, mainly relied on surveys and questionnaires, and was carried out by individual researchers rather than by institutional or government funding.

With the 2000s, new urban spaces were created. As a result, the behavioural and interactional patterns of the youth in digitally mediated urban sphere drastically shaped the agenda of youth studies. The research became more cross-disciplinary and exhibited methodological pluralism. Studies explored the changing social and cultural habits of the youth (e.g., Özensel, 2009; Yazıcı, 2001), youth participation through online channels (e.g., Neyzi, 2011; Telli-Aydemir, 2009), online youth cultures (e.g., Tuzcu Tığlı, 2019), emerging identities and gender (e.g., Alemdaroğlu, 2007; 2010; Çelik & Lüküslü, 2010; Demez, 2009; Yonucu, 2005) youth unemployment (e.g., Yentürk & Başlevent, 2008; Yücel & Lüküslü, 2013), youth subcultures (e.g., Semerci, Erdoğan & Sandal Önal, 2017; Şişman, 2013), and most recently climate activism and youth (e.g., Atik & Doğan, 2019).

Neyzi (2001) argues that public discourses constructed around youth have changed in accordance with the social, political, economic developments in Turkey. She underlines that between 1923-50, the youth were conceptualized as the nation itself, they were heroes and the future of the nation. Later in the 1950-80 period, youth were reconceptualized as threats and rebels. In post-1980, the plurality of labels manifested and the youth started to construct and challenge the discourses which label them. In her doctoral dissertation exploring the perceptions and definitions of 15-24 year-olds regarding the conceptualization of the term youth in Turkey, Demir (2021, p.97) underlines a similar point by indicating that youth studies in Turkey should refrain from defining the youth but rather give an ear to them. Until 2000s, the youth research in Turkey failed to acknowledge that the young people are entitled to construct or express their own discourses. The current agenda on youth studies, though, utilizes various solid methodologies across different scholarly fields in Turkey. The research on linguistics with a focus on youth, in this sense, has the potential to provide an elaborated emic perspective which long have been ignored in the informing fields.

2.4.2 Linguistics

Existing linguistic studies of Turkish youth talk are scarce and they are oriented towards the variationist paradigm. Overall, these studies approach the linguistic repertoire of young speakers, which consists of high school and university students, as a divergence from the so-called standard variety of associated with the wider society.

The earliest study to date is Açıkalın's (1991) work which indicates that 17-19 year-old speakers of Turkish deviate from the language they speak at home and the standard language when they speak among their peer group. The study argues that each generation has a different language usage. The study lists the motivations behind using a language exclusive to a group as the desire to be part of a community, expressing intense emotions and experiences, and keeping up with the social developments. Even though there is limited information regarding the profile of participants and the scope of data collected, the notable thing about the study is that it makes use of naturally occurring data and provides thick description for the data presented.

Remaining studies make use of highly structured and elicited forms of data. The methodological tools reported in these studies are limited to close-ended

questionnaires, observation notes, structured interviews and document analysis. While these studies highlight that language is dynamic and interactional in nature, naturalistic data is not the focus of their queries. Toğrol (2012), for instance, lists the frequently used adjectives used by Turkish male speakers between the ages of 13 and 17 based on the data collected via a questionnaire. Similarly, Şafak and Bilginsoy (2019) investigate the use of neologisms derived from a pre-determined list and explores their frequency of usage among what the researcher calls "Turkish Z Generation" through a Likert-scale questionnaire. What is notable about the study is that the second author was a high school student. A total number of 50 items were developed based on the fieldnotes of the researchers and the questionnaire was administered to 100 students from 5 high schools in the province of Kırklareli. The results indicate that young people tend to use the words in their original language, i.e., English, rather than Turkish equivalents provided by Turkish Language Association.

Apart from these studies which takes a generationist perspective and a relatively quantitative approach, there are also a few researchers with prescriptivist orientations who advocate for the argument that youth language is a non-standard, deficit, transitive form of language which is a divergence from the proper usage of Turkish (Canbulat, 2017; Gunay, 2007). Canbulat's work (2017) focuses on views regarding youth talk elicited from students enrolled in a Turkish Education Department. The study reports that majority of participants were in consensus that the youth language corrupts Turkish language. The participants also provided reasons behind using youth language as positioning 'different and cool', being an in-group member, alienating the out-group, and facilitating communication within the group. Within the scope of discussions regarding lexical borrowings in Turkish, Gunay (2007) dubs the phenomenon as "trendy borrowings" and points at the youth as the perpetrator. The use of lexical borrowings in youth talk is argued to be appearing as "intellectual, stylish, and flamboyant" (2007, p. 51). Both of these studies view the language practices of youth as a transient phase that needs to be completed (Canbulat et al., 2017).

The research agenda also includes discussions regarding the influence of new media tools over the linguistic practices of young speakers of Turkish. Kırık (2012) briefly argues that social media has 'negative' effects on Turkish youth language while Çokol (2020) mentions 'generational conflicts' experienced due to deviations in language spoken by different age groups in Turkey and provides personal observations and

examples of language produced by young speakers in social media. Öztürk Dağabakan (2017) compares Turkish and German address terms used by youth in social media and lists the use of abbreviations, omissions, vowel/syllable insertions to the words, and phonetic changes as common characteristics of written social media discourse in both languages. These studies are accounts of observations with a focus on language change observed in Turkish yet they fail to offer a systematic and empirical investigation of the language.

As opposed to the bulk of existing studies, Demir (2010) opposes the myth that youth language corrupts the standard Turkish and that it has a limited lexicon. The study briefly mentions that specific words and expressions are observed more frequently in youth language and labels it as a generation-specific way of speech.

At this point, it is worth mentioning the studies conducted within the intersection of fields of lexicography and Turkish language teaching. Apart from Aksan and Uçar's (2012) and Kurtoğlu and Uçar's (2011) methodologically guiding works which underline the need for utilizing corpus tools for data driven learning and teaching, a large part of studies focus on describing the core vocabulary in written language produced by students in different grade levels and then either explore the influence of gender, age, socioeconomic status on size and composition of the identified sample of lexicon or suggest guiding principles or models for teaching language skills -particularly vocabulary teaching and reading comprehension- in Turkish (Aksoy, 1936; Büyükkantarcıoğlu, 1992; Çıplak, 2005; Davaslıgil, 1980; Harıt, 1971; İpekçi, 2005; İpek Eğilmez, 2010; Karadağ, 2005; Kurudayıoğlu, 2005; Pars & Pars, 1954; Temur, 2006; Tosunoğlu, 1988; Yazı, 2005). While the participants in aforementioned studies consist of students in primary education in Turkey, there are a few studies which explored written data produced by high schoolers (Koçak, 1999) and university students (Çiftçi, 1991; Pilav, 2008). Research investigating the lexicon of spoken language produced by Turkish speaking students is even more scarce. So far, the studies have solely made use of data elicited from structured or semi-structured interviews which lasted for 5 to 10 minutes for each student (Emiroğlu, 2015; Obuz, 2012; Ünsal, 2005). Similar to the studies focusing on written language, studies on spoken language depicted the types, tokens, frequencies of words used by speakers and presented the distribution of data with regard to different demographic parameters such as gender, grade level, socio-economic status in their samples.

Additionally, there is also another group of studies which either propose a core spoken or written vocabulary to be used in teaching Turkish as a foreign language to youth or explore the scope of vocabulary already present in teaching materials designed for such learners of Turkish as a foreign language (Arslan & Durukan, 2014; Aşık, 2007; Bozkurt, 2015; Bulundu, 2022; Göçen & Okur, 2016; Hayran, 2019; Tunçel, 2011; Yahşi, 2020). In this sense, it can be argued that these studies primarily contribute to the discussions concerning Turkish language education practices in both macro level (i.e., national language planning and education policies) and micro level (i.e., teacher practices and material selection/development) in Turkey. The underlying scholarly motivation in these studies, then, has been to suggest a representative vocabulary adjusted for cognitive levels or learning goals of speakers of Turkish. This study, on the other hand, does not problematize the language spoken by a specific speech community (i.e., the youth) but rather approaches the data from a socio-pragmatic perspective and explores the linguistic devices, patterns and strategies observed in youth language via corpus methods.

In this chapter a review of the related literature with regard to youth language and corpus linguistics was provided. In the following chapter, the method of the study will be presented.

CHAPTER 3

METHOD OF RESEARCH

3.0 Presentation

This chapter aims to inform the reader about the research design of the study, the process and the characteristics of the specialized spoken corpus constructed, methods and tools of analysis used, and the issues of reliability, validity and ethics. Firstly, the research design which is shaped by the research questions and three important building blocks which guided corpus construction (authenticity, representativeness, size) are presented. Then, the sampling criterion for determining the sample, the procedures for recruiting participants and the process of informing the participants are introduced in detail. The data collection tools, namely the semi-structured interviews, audio recordings, and the demographic information form are presented and the timeframe of the research is explained. The piloting stage and the data collected are also presented. After introducing the corpus construction and analysis software EXMARaLDA, the parameters for choosing the software will be explained. Three components of the software will be introduced (Partitur-Editor, COMA, and EXAKT) and the workflow of corpus construction process will be presented. The components of metadata of the constructed corpus; the communication metadata and speaker metadata are outlined. Transcription conventions and annotation scheme are presented. As for the presentation of how data analysis was conducted, the implementation of corpus methods into discourse analysis and the affordances of integrating three main corpus analytical methods, namely frequency lists, the KWIC analysis, co-occurrence are explained. Following the choices regarding tools and procedures of analysis, the concerns regarding reliability and validity as well as the ethical considerations are presented.

3.1 Research Design

This section has two parts. In the first part, the research questions are presented. In the second part, three important tenets of corpus construction: authenticity,

representativeness, and size which monitored the process of corpus design are explained.

3.1.1 Research Questions

Guided by the purpose of compiling the first corpus of Turkish youth language, this study aims to examine the linguistic characteristics and discursive dynamics of dyadic and multi-party youth interaction in contemporary spoken Turkish. In this vein, the study addresses the following research questions under two complementary layers:

Layer One: Corpus Construction

1. What is the structural composition of the Corpus of Turkish Youth Language (CoTY)?

- **a.** How many tokens and types does the corpus encompass?
- **b.** What is the distribution demographics for speakers in the corpus?
- **c.** What is the distribution of data with regard to speakers and speaker groups?

Layer Two: Linguistic Architecture

2. What are the typical *topical* and *lexical* characteristics of the interaction among young people between the ages of 14-18 in the CoTY?

- **a.** What are the dominant topics and sub-topics observed in the corpus?
- **b.** What are the key concepts and typical vocabulary identified for the corpus?

3. What are the *interactional markers* used by young people between the ages of 14-18 in the CoTY?

- **a.** What are the *response tokens*, their types, frequencies and functions?
- **b.** What are the *vocatives* their types, frequencies and functions?
- **c.** What are the *vague expressions*, their types, frequencies and functions?
- **d.** What are the *intensifiers*, their types, frequencies and functions?

These complementary layers are illustrated in Figure 1 below.

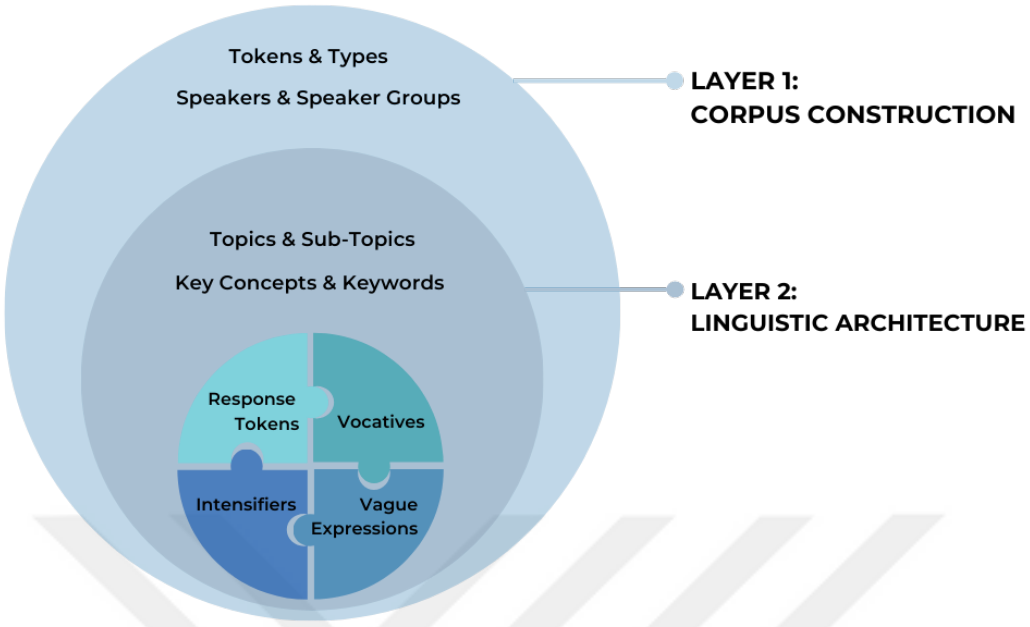


Figure 1 Foci of investigation in each research question layer in the study

In order to answer the aforementioned research questions, this study has the ultimate aim of building a specialized corpus as a source of data and as a tool of analysis. For this purpose, The Corpus of Turkish Youth Language (CoTY) was designed, compiled, and constructed. The corpus was designed to encompass various modes and mediums of youth interaction and expand over the years, yet this dissertation exclusively focuses on spoken data.

The general consensus for the defining characteristics of a corpus is that it consists of (i) machine-readable, (ii) authentic texts, which are (iii) sampled to be (iv) representative of a particular language or language variety (McEnery et al., 2006, p. 5). While machine-readability is the de facto characteristic of modern-day corpora, the qualities of authenticity and representativeness need further elaboration. These concerns, along with the issue of size, will be explained in relation to the CoTY in the following sub sections.

3.1.2 Authenticity

Sinclair (1996) discusses the quality of *authenticity* of a corpus by highlighting the need for 'minimum disruption' regarding the ways the linguistic evidence is collected. If data collection is unobtrusive to the communication between individuals whose data is gathered, the language behaviour is less expected to be distorted and the data can be defined as relatively more authentic. In more broad terms, authenticity is described as the 'the real-life language use' by McEnery and Wilson (2001). It should be noted that authenticity is a relative term and it includes numerous aspects of the data collection procedure and data itself. Therefore, the contextual information regarding the communication, the speakers, the setting, as well as the ways and details of representation of this information have direct consequences to the degree of authenticity of a corpus.

In this study, even though all the participants in the study were informed beforehand regarding the audio recordings, several measures were integrated in order to minimize the disruption and increase the authenticity. First of all, this study was designed in a way that the data was collected in informal settings and without the presence of the researcher. As will be explained in more detail in section 3.2 of this chapter, an in-group member recorded the conversations. Additionally, rather than a recording device, built-in audio recording capabilities of mobile phones or personal computers were utilized. In this way, neither the participants nor the setting was interfered during data collection. Secondly, the speakers were not manipulated in any way regarding what to talk about, they were reminded to chat as they always did. There were no restrictions to 'the type of language, expressions, or words used' and the 'topics' mentioned in their talk. Underlining that their private information would remain anonymous contributed to increasing the degree of authenticity of the data, as well. Even though all of the participants knew that they would be recorded, the researcher asked the in-group member, the informant, to refrain from informing the exact time of the recording to other participants if possible. Thirdly, the researcher conducted post-interviews with a random sample of participants to ask them to evaluate their own level of comfort and naturalness during the conversation. Additionally, the informant who was responsible for providing the metadata of the conversation and the speakers was asked to note down additional comments regarding any unusual or notable aspect about the conversation in

the Recording Log (please refer to section 3.3 for more information on data collection tools, and Appendix E for the full sample log).

3.1.3 Representativeness

Representativeness is regarded as a fundamental quality of designing a reliable corpus and it addresses the issue of the representation of parameters such as the type of materials, speakers, language varieties a corpus aims to represent. In other words, the concept proposes that there must be a match between the language being examined and the type of material the corpus contains (Biber, 1993). The notion of representativeness is multifaceted and is directly related to research questions a corpus seeks to answer. A requisite in corpus construction is that the sample in a corpus should be representative of the speech community in question. To elaborate, the representativeness needs to be ensured so that 'a group of cases taken from a population that will, hopefully, represent that population such that findings from the sample can be generalised to the population' (McEnery & Hardie 2012, p. 250).

The notion of *balance* is directly related to the notion of representativeness, as well. A corpus is regarded balanced if it covers 'a wide range of text categories which are supposed to be representative of the language or language variety under consideration' (McEnery et al., 2006, p.16). In other words, if the size of sub-corpora representing particular genres or registers in a corpus is proportional to the relative frequency of occurrence of those genres in the language's textual universe as a whole, it is considered as a balanced corpus. (Leech, 2007, p. 136).

Given the multiplicity of issues regarding achieving representativeness and balance (e.g., determining all features of a language/variety and actual distributions of all the potentially relevant parameters in any population to devise a sampling frame), in majority of cases it is inevitable that the corpora will be relatively skewed if they are to be compared to the overall population in the end. That being said, as the notion of representativeness is criticized to be a rather vague term by Sinclair (2005) and it is already acknowledged that the ideal level of representativeness is practically not possible to attain (Stefanowitsch, 2020; Adolphs & Knight, 2010), the general consensus in corpus linguistics is that the concepts of representativeness and balance can be used

as guides to design the overall structure of the corpus in order to achieve a 'reasonable representation' (Kilgariff et al., 2006, p. 129).

For this study, several approximations were carried out to capture a maximally representative sample of the population in question. First of all, the boundaries of the linguistic data to be collected were clearly defined. As the CoTY is a specialized corpus, the data belongs to a single register: spoken language. The mode of communication has two parameters: face-to-face or online interaction. The production of speech is spontaneous and naturally occurring. The immediate situation is limited to informal settings which consist of both indoors (e.g., bedroom, kitchen, living room) or outdoors (e.g., street, park, backyard of house) and main communicative purpose is defined as personal communication among friends. As it is informal communication between friends, the audience domain is characterized as private and the participants have symmetrical relationships. Table 1 below summarizes these register characteristics.

Table 1 Register characteristics of the CoTY

Mode*	***Spoken: face-to-face and online***
Interactiveness and production	Spontaneous with a degree of advanced planning
Shared immediate situation	Informal settings: indoors, outdoors, virtual
Main communicative purpose	Personal communication
Audience	Private
Participant roles	Symmetrical: friends

*based on Biber et al., 1999, p. 15-17

Apart from these register characteristics which are stable within the whole data of the corpus, the parameters of sex and provinces are balanced in accordance with the overall population they represent. As a result, a more principled approach to representativeness is adopted. As the CoTY aims to be maximally representative of the youth language spoken in Turkey[8], the relevant proportions of the sample are designed based on the statistics reported by Turkish Statistical Institute (Tur. *TÜİK: Türkiye İstatistik Kurumu*) which is the official government agency commissioned with compiling and producing

[8] At this point it is important to note that the speakers in the CoTY consist of young people who are enrolled in national education system in Turkey, the young people who are NEETs or are in active labour market but not in education are not represented in the sample.

national statistics in relation to population, economy, environment, culture and other related areas. According to Youth in Statistics 2021[9] by Turkish Statistical Institute published in 2022, the designated cohort of 15-24 years which is defined as 'youth' by the institute corresponds to 15.3% of the whole population in Turkey. As Table 2 below shows, males make up 51.2% while females make up 48.8% of this age cohort and the provinces which has the highest number of youth population are Istanbul, Ankara, and Izmir, respectively.

Table 2 Youth population by province and sex in Turkey

Province*	***Youth (15-24 years) in Turkey***		
	Population	***Males***	***Females***
all (country)	12,971,289 (100%)	6,648,929 (51.2%)	6,322,360 (48.8%)
İstanbul	2,339,946	1,199,887	1,140,059
Ankara	846,595	433,071	413,524
İzmir	572,286	297,055	275,231
Bursa	428,251	220,833	207,418
Şanlıurfa	401,523	202,883	198,640
Konya	375,916	190,103	185,813
Gaziantep	368,26	187,084	181,176
Antalya	356,151	182,79	173,361
Adana	338,841	173,393	165,448
Diyarbakır	325,599	166,157	159,442
Kocaeli	291,268	151,034	140,234
Mersin	273,382	139,502	133,88
Hatay	264,31	136,522	127,788
Van	231,211	118,907	112,304
Kayseri	226,602	116,693	109,909

*first 15 provinces out of 81 in the country are listed here.

As the sample of this study is designed to maximally represent the target population, the proportions of sex of the participants and the provinces participants reside in are designed to correspond to the proportions in overall population. At its current scope, the CoTY does not aim to make any linguistic generalizations for the whole 15-24 age cohort in Turkey but rather explore the linguistic evidence systematically collected through transparent and consistent sampling parameters which are based on the overall

[9] Please visit https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Genclik-2020-37242 for detailed information for the statistics regarding 15-24 age cohort in Turkey.

characteristics of the greater sample. The parameters such as the sex ratio and the proportions of provinces are used as guides to adhere for compiling the corpus. In this way, the scope of the CoTY can be expanded in the future by means of increasing its degree of representativeness, rather than compromising it.

3.1.4 Size

One of the issues regarding the corpus design is to determine the *size* of corpus to be compiled. While the initial trends favoured *mega-corpora* (e.g., Bank of English which is a written corpus with 650 million running words), the recent turn in corpus linguistics has started to highlight the advantages of smaller specialized corpora which provide deeper insights into the contextual features of the linguistic patterns observed within the sample. As one size does not fit all, rather than trying to determine the corpus size a priori in the design stage, the appropriate size of a corpus depends on the aims of a particular research and is finalized after cyclical turns of data collection and monitoring of the parameters for achieving reasonable levels of representativeness.

Though Sinclair (2004, p. 189) famously states that a corpus should ideally be big as 'small is […] simply a limitation', the recent bulk of research underline that a corpus does not need to be as large as a general corpus to yield reliable results (Biber, 1990; Carter & McCarthy, 1995; Egbert, et al. 2022; Flowerdew 2004; Koester, 2010; Reppen, 2010; Tribble, 2002) but rather it should ensure a minimal sample size which is directly related to the previously explained notion of representativeness. In terms of the number of samples required, Biber's (1990) work in which he tested the number of text samples minimally required to represent a register/genre in a corpus is noteworthy to mention concerning the discussions of corpus size. His statistical analysis showed that linguistic tendencies are quite stable with ten text samples per register/genre and the most common linguistic features are relatively stable in their occurrence across 1,000-word samples. Biber's work is significant in the sense that it certifies that it is not mandatory to build a mega-corpus of millions of words to yield reliable results. It is also important to note that in case of a spoken corpora, controlling the length of the samples (i.e., the transcripts of audio recordings) is challenging, thus the number or the type of registers/genres a spoken corpus encompasses are generally more focused compared to the design of written corpora. While the spoken corpora projects complied by means of funding opportunities can be relatively large in size (e.g., The Spoken BNC2014 which

has 11.5 million-words compared to its written counterpart which has over 100 million-words), a big number of spoken corpora are smaller in size and constructed with a more focused aim of linguistic investigation. Similarly, the COLT and the COLA-m, which can be considered as the predecessors of youth language corpora, are relatively modest in size, consisting of 444,166-words and 463,047-words, respectively.

This study focuses on a single spoken register of informal youth talk among friends in order to investigate linguistic and discursive characteristics of this particular register within the scope of its sample. As will be explained in more detail in the following section 3.2.1, the sampling frame was designed to maximize the amount of the data to be collected.

3.2 Participants

In this section, the sampling frame, methodology adopted for recruiting participants, and roles of the participants will be presented.

3.2.1 The Sampling Frame

Several parameters were set for selecting the participants to be included in the sample of the corpus. Defined sampling frame includes recordings of interactions which are (i) naturally occurring and spontaneous speech, (ii) in informal contexts, (iii) between speakers who are friends, (iv) currently enrolled in high school or graduated and studying for the university entrance exam, and (v) whose native language is Turkish.

In order to address the issues of representativeness and balance, data collection was completed in two batches. In 'the first batch', the participants were contacted through convenience sampling; the researcher approached the target participants in accordance with the selection criteria. The participants were trained for data collection and were assigned a timeframe to submit recordings to the researcher. The first batch of recordings acted as the point of reference for the researcher to identify the imbalances in the sample and arrange 'the second batch' of participants who were complementary to the first batch so that the balance between sex of speakers and the number of participants in each grade level were assorted. In other words, the sampling strategy for the study started with convenience sampling in which snowball sampling was embedded

and was followed by maximal variation sampling. Similarly in the construction of The Spoken BNC2014; Love, Dembry, Hardie, Brezina, and McEnery (2017) adopted a similar approach which they call the *opportunistic approach* to data collection in which they targeted recruiting specific groups of people via advertisement campaigns in social media, students' recruitment campaigns at universities and press releases when imbalances in the data appeared.

As explained earlier in previous sections which dealt with the representativeness and corpus size, the participants in this project were sampled in order to maximally represent the population in question. The ratio of sex was set similar to that of the overall population and the distribution of provinces (both in terms of cities of residence and hometowns) are guided by the proportions in the overall population. Similar to the overall youth population in Turkey, the highest number of participants are from the provinces of İstanbul, Ankara, and İzmir in the corpus as well. Additionally, the proportions between the grade levels were designed to be as balanced as possible. The sample in this study consists of five sub categories for grade levels which are high school grades of 9^{th}, 10^{th}, 11^{th}, 12^{th}, along with an additional category for graduate students who completed high school education but had not started university yet. For each grade level, the number of students and the sex is designed to be maximally balanced.

As with all corpus building projects, a number of compromises were made within the sampling frame. A number of parameters, namely the socio-economic status, school types, and the cities of residence were not controlled during data collection so as to maximize the number of participants. Though not controlled during the data collection stage, the results show that socio-economic profile of the participants within the corpus is relatively balanced. In terms of other parameters, the sample shows variety, as expected. Detailed information regarding the overall profile of the participants based on the sampling parameters are presented in Chapter Four.

3.2.2 Recruiting Participants

As briefly mentioned in the previous section, participants were invited to the study through a variety of recruitment methods. Individuals were approached directly by the researcher, by the participants themselves, and also by gatekeepers who had access to specific groups of young people.

The researcher promoted the aims and the scope of the study via an online flyer (Appendix A) and a website and contacted participants who volunteered for the study. The families/guardians of the participants were either personally visited or were phoned to explain the purpose, the design, and the procedure of the study and answer questions if there were any. If the family/guardians and their children agreed to cooperate with the researcher for the study, the informed consents of both the participant and their parent/guardian (Appendices B-1 & B-2) were collected. The families/guardians were also informed that the study required children to do recordings with their peers, therefore the consents of the families/guardians of those peers were taken as well. The participants were also asked to recruit new participants by promoting the study among their peer groups. The researcher also promoted the study online, and briefed gatekeepers of several online youth groups who showed interest for participation. The gatekeepers conveyed the invitation for participation to their own audience and individuals who had additional questions contacted the researcher directly.

3.2.3 Roles of Participants

The recent focus in public engagement in science have led to the emergence of a *participatory turn* (Jasanoff, 2003) in various fields of scholarly research. Citizen sociolinguistics proposes a shift from the focus on traditional researcher role which has the authority over the knowledge to a stronger emphasis on participants as the legitimate holders of knowledge about the language data (Rymes & Leone, 2014; Svendsen, 2018). The importance of inclusion of emic perspective is highlighted in the studies. In this vein, this study utilizes contributory public participation model (Shirk, et al., 2012) through which the degrees of data precision and accuracy are increased.

Contributory public participation model informed the participant recruitment process and the roles of participants as well as the data collection process. Within this scope, participants had four vital roles in the study: (i) recording the conversation with their peers and submitting it to the researcher, (ii) providing demographic information and detailed metadata about all the speakers in the conversation by filling in a Recording Log, (iii) introducing the study to the new potential participants, (iv) acting as informants to provide an emic perspective for researcher to identify unclear or unintelligible

utterances encountered in the conversation, as well as the validation of the interpretations if needed.

The researcher met the participants (face-to-face or through online channels), explained the study, their roles and briefed them about how they needed to proceed with the data collection process. Each participant was also handed a short guide as an overview of important technical and contextual reminders for data collection. The researcher provided the participants with her contact details in case the participants, potential participants or parent/guardians had further questions.

3.3 Data Sources and Data Collection Timeframe

Data collection timeframe was designed to first carry out a piloting stage to test the data collection tools and then revise the tools and collect the data for the main study. The piloting stage of this study took place in September 2019 (see 3.4 for detailed information on this stage). The main data collection procedure was scheduled to start in October 2019 and last for six months. Nevertheless in March 2020, the first official case of COVID-19 pandemic was reported by the Ministry of Health in Turkey and a series of measures was implemented to limit the social and physical interaction among people. As a result, the data collection came to a halt due to the unfolding events in relation to the unforeseen outbreak. In the following weeks, all of the initially recruited participants dropped out of the study. Given the new circumstances with regard to the pandemic, the designated timeframe for data collection was revised to take place between November 2020 to October 2021 and the data were collected both face-to-face and via online means (see Appendix C for the detailed timeline for data collection period).

The main data sources for this study is naturally occurring spontaneous speech data which was collected by means of audio recordings. The in-depth demographic information about the speakers and context were collected through another data source in the form of a questionnaire which is called Recording Log. These two data sources are complementary for the process of construction of the corpus. Another data source is semi-structured interviews which were particularly utilized for piloting stage and the results provided researcher with a preliminary description of the profile of the group of people under investigation. Each of the data sources will be presented in following sub-sections.

3.3.1 Interviews

The interviews were designed as semi-structured (Appendix D) which were conducted either face-to-face, or via online means (i.e., Skype or WhatsApp Video Call). Interviews allowed the researcher to introduce the study to individuals in detail and in the meantime obtain information on the daily and online routines, activity types, personal characteristics and interests of the individuals as well. Based on the questions posed and the needs of the participants, the researcher was able to revise the procedure or instructions based on the questions and needs of the participants. Interviews were utilized as a data collection tool for the piloting stage, thus more detailed information will be presented in section 3.4 of this chapter.

The interviews were scheduled according to the agenda of the participants, were conducted in Turkish and audio recorded. A *gisted transcription* was carried out and the essence transcript format (Dempster & Woods, 2011) was used in order to capture the highlights of the interview content. Essence transcripts are selective in the sense that the parts of data relevant to the research purpose of the study are represented. Through this procedure, the researcher was able to summarize the profile of the participants along with her comments regarding the content of the interview data.

3.3.2 Audio Data

The initial, or rather the first generation, spoken language corpora included transcriptions of speech without any access to audio data (e.g., the BNC, London-Lund Corpus) where it was not possible to carry out investigations of pauses, silences, paralinguistic and prosodic features. Later the advancements in technology introduced the modern equipment for data recording such as high-quality digital voice and video recorders which enabled researchers to collect richer data in a more unobtrusive way. The existing spoken corpora such as the Spoken component of The British National Corpus (The Spoken BNC1994) and The Spoken Turkish Corpus (STC) supplied the recruits with voice recorders to collect data. In both the BNC and the STC, recruits who recorded the conversations logged details of each conversation after the recordings (in a special notebook for the BNC and in recording information sheets for the STC). Later in more recent spoken corpora construction, audio recording capabilities of smart phones was utilized (e.g., the BNC2014).

This study made use of both smart phones (in face-to-face or online interaction) and computers (in online interaction) to collect the data. The informant in each group of speakers was responsible to use their own smartphones or online communication platforms (e.g., Skype, Zoom, etc.) to record the conversation and then submit the recording to the researcher. Digital turn in data collection made data collection faster and briefing the informants easier.

The briefing of the informants was conducted in a form of a short meeting which involved conveying technical instructions and contextual reminders. Below is the overview of technical highlights communicated to the informants:

- Before you start the recording, confirm the default location of the recordings. Check whether you have enough storage space available for the recording.
- Make a trial recording of a few seconds. Find the location of your recording and check the recording for voice quality.
- Refrain from over-crowded places or settings where there are high levels or obtrusive types of background noise that might block the conversation.

For the contextual reminders, the informants were reminded that:

- There are no restrictions in terms of topics of talk.
- You can use slang, swear words, foreign words and expressions. Just speak naturally as you usually do.
- All the proper names or any other private information which may identify you or any person mentioned in the recordings (e.g., your name, name of any person mentioned, name of your school/neighbourhood, phone numbers, email addresses, etc.) will be anonymized.
- There is no limit to the maximum duration of the conversation. You can record a single conversation in a single session or do successive recordings.
- After you complete the recording, do not forget to fill in the Recording Log. Fill in a form for each recording you completed.

All the data were converted to .wav format, assigned unique IDs, classified and stored electronically.

3.3.3 Recording Log

Recording Log designed for this study is a questionnaire which consists of both close-ended and open-ended questions with regard to the metadata about the speakers, the setting, and the interaction as a whole. The structure, the wording and the sequence of the log was piloted before administering it to the informants. Following the piloting, the log was also reformatted as an online Google Form sheet so that it would be time-efficient for both the informants and the researcher. Using an online form also enabled the researcher to save time to obtain the data and to monitor the accumulating metadata synchronously.

Recording Log (see Appendix E) consists of 7 sections. In the first section, the contact details of the informant, the date, the duration and the name of the recording are filled in. In sections 2, 3, and 4, demographic information about the speakers and their families are recorded. Each section is designed to collect data of an individual speaker in the conversation, therefore sections 2 and 3 are compulsory while 4 remains optional. Section 5 describes the context (the place where and when the conversation took place) for each recording. Section 6 requires information concerning the relationship between the speakers (when and how the speakers met, the frequency of communication between the speakers). Lastly in section 7, the informants are invited to provide optional commentaries regarding the recording if needed.

The log aims to obtain as much metadata about the speakers and the context as possible so that the spoken language can be discursively interpreted for the identified linguistic practices.

3.4 Piloting

In order to test the data collection tools and the planned procedure for data collection, firstly a pilot study was designed and carried out. The piloting consisted of three parts: semi-structured interviews, audio recording collection, and respondent feedback sessions, respectively. Total number of participants in this stage consisted of 10 individuals (5 males, 5 females) between the ages 14-18 enrolled in high schools in Ankara, the capital of Turkey. A top-down approach was adopted for the selection of the school types in order to ensure a diversification in the socio-economic status of the

participants. As previously mentioned, the main study did not have school type and socio-economic status of the families as controlled parameters in the participant selection criteria for main data collection. Therefore, the profile of the participants in the piloting was designed in a way that they can respond to the potential diversification within school types and socio-economic status parameters in the main study.

The studies on socio-economic development levels of the districts of Ankara (Demir, 2017; Dinçer & Özaslan, 2004; Özaslan, et al., 2006) were consulted to choose the schools where the pilot data were to be collected and diversify the socio-economic background of the participants.

Below in Table 3, the profile of participants in pilot study is presented.

Table 3 Pilot study participants

Pseudonym	***Sex***	***Grade Level***	***School Type***	***Location***
Ahmet	M	9	Anatolian High School	Kızılay
Bora	F	10	Science High School	Şentepe
Berrin	F	10	Anatolian High School	Mamak
Cansu	F	10	Anatolian High School	Etimesgut
Ceyhun	M	10	Private High School	Çayyolu
Berk	M	11	Anatolian High School	Batıkent
Hande	F	11	Vocational High School	Keçiören
Burak	M	11	Private High School	Söğütözü
Serhat	M	12	Vocational High School	Kızılay
Gamze	F	12	Private High School	Batıkent

In the pilot study, it was ensured that there was at least one participant in each of the grade levels and that school types showed variety. The pilot study included private high schools (*n*=4), Anatolian high schools (*n*=3), vocational high schools (*n*=2), and a science high school (*n*=1). In terms of location, the schools are scattered across 5 districts of Ankara, namely Çankaya (Kızılay, Söğütözü, Çayyolu), Yenimahalle (Batıkent, Şentepe), Mamak, Etimesgut, and Keçiören. According to Demir's (2017) investigation of socio-economic development levels of the districts of Ankara; Çankaya is the most developed district out of 25 districts in Ankara, while Yenimahalle placed 3rd, Etimesgut 4th, Keçiören 6th, and Mamak 9th place in the list.

The piloting took place in September 2019, the interviews and due dates for the submission of the collected data were scheduled according to the agenda of the participants.

In semi-structured interviews, the researcher obtained information about the daily routines (online and offline), the interests, aspirations and dreams, relationships with friends, and personal characteristics of the participants. The interviews were conducted either face-to-face or via online means (e.g., Skype) depending on the preference of the participant. Interviews lasted between 9 to 20 minutes each and made up 3 hours and 30 minutes in total.

Semi-structured interviews in the piloting phase revealed that the participants spent most of their spare time with their friends from school and they also tended to spend their time in cafes and parks closer to their schools. Therefore, the profile of the school location can be regarded as a prominent social factor affecting the sociocultural and linguistic conceptualizations of the participants. The interviews also indicated that the participants used social media quite frequently both as a means of connecting with their friends and also to get engaged with their interests and hobbies. All of them reported using photo sharing application and social network platform Instagram, three of the participants also had Twitter. In addition to Twitter, two of them also had multimedia messaging application Snapchat. The participants indicated that they did not watch television, they watched 'YouTube' or 'Netflix'. Those who were in their last two years of high school reported to spend most of their time studying and preparing for the university entrance exam. All of the participants shared the dream of 'going abroad', be it for travel, study or work.

At the end of the interviews, the participants were briefed about the data collection process and their responsibilities. The summary of the highlights regarding these issues were also provided as a separate document for them to consult if needed. The participants were required to collect data from their peer groups in informal settings, the number of speakers would be maximum 4 people and the duration of the recordings would be minimum 15 minutes. Each participant was asked to do a test recording and send it to the researcher before their actual data collection in order to avoid technical problems. The participants also had run through the Recording Log with the researcher in order to check whether there was any item they did not comprehend. The researcher

asked the informants to reflect back on their tasks to check whether there was any point missing or misunderstanding. The participants were asked to fill in the Recording Log after the conversation and collect consent forms before they did the recordings. The researcher provided the participants with her contact details and the website of the project in case a participant or parent/guardian of the participant requested additional clarification concerning the study.

After the allocated time for audio data collection period of the pilot study was completed, the participants submitted 16 recordings which corresponded to 57 minutes in total. Following the submission of recordings, the researcher contacted the participants to inquire about their experience regarding the procedure and reflect on the audio recordings as the last step of the pilot study. Based on these correspondences, several revisions were made in the data collection procedure and tools. The revisions are as follows:

- Participants were encouraged to record conversations around 20 minutes to obtain more contextual cues regarding the topics and the interactional goals identified in the conversation.
- The number of speakers was determined to be maximum 3 people to decipher the overlaps in talk and minimize problems regarding speaker identification.
- Metadata in the Recording Log was expanded. Additional parameters of education level of parents, occupations of parents, frequency of communication, grade point average for the current school year were integrated in order to obtain a more detailed profile of participants and a thicker description of the interaction.
- Consent Forms were reformatted as online forms.

The audio data of the pilot study were transcribed orthographically. Later, transcription conventions and annotation scheme to be used for the main study were piloted to test whether they match with the needs of the study and also whether they suit the inherent characteristics of the youth language data. This stage led to the development of several additional annotations for paralinguistic and prosodic features which are required to be tagged in the data (See section 3.7 of this chapter for transcription conventions and the annotation scheme used in the main study.)

3.5 Corpus Building and Analysis Software Used: EXMARaLDA

In order to select the corpus software to be used in the study, a selection criteria was set based on the purpose, the research questions, and the inherent nature of the data and the planned analyses of the study. The selection criteria for a multilayer transcription and annotation software programme was determined as follows:

- The corpus tools compatible with handling spoken language data
- Turkish keyboard support
- Time-aligned transcription of audio
- Flexible annotation options
- Detailed metadata for the speakers and transcription files
- Integrated corpus query and analysis tools
- XML based data format
- Flexibility for output formats
- User-friendly interface
- Availability of technical support

Based on these criteria; the software programmes ANVIL, CLAN, ELAN, EXMARaLDA, and TRANSANA, all of which are developed for analysing naturally occurring interaction data were selected as candidate softwares that could be used. Affordances and limitations for each software were compared to determine the most suitable one for the needs of the study. Below in Table 4, the overview of their affordances and limitations are listed.

Table 4 Comparison of multilayer transcription software programmes

Feature	*ANVIL*	*CLAN*	*ELAN*	*EXMARaLDA*	*TRANSANA*
frequently used for	gesture research	language acquisition research	psycholinguistics	conversation and discourse analysis	conversation analysis
time-aligned transcript and media	YES	YES	YES	YES	YES
speaker metadata	YES	YES	YES	YES	YES
flexible annotation	YES	YES	YES	YES	YES

Table 4 (cont'd)

Feature	*ANVIL*	*CLAN*	*ELAN*	*EXMARaLDA*	*TRANSANA*
scope of query	basic text search only	word search, frequency, mean length of utterance	basic text & code retrieval	complex queries	keyword search & frequency analysis
open source	YES	YES	YES	YES	NO
multilevel annotations for a single tier	YES	YES	NO	YES	YES
handling overlapping speech	YES	YES	YES	YES	YES
transcription capacity	limited	turn-by-turn	partition	partition	turn-by-turn

As EXMARaLDA proved to address all of the prerequisites and highest number of affordances among the available options, it was selected as the software to be used for this study.

EXMARaLDA (*Extensible Markup Language for Discourse Annotation)* software (Schmidt & Wörner, 2014) consists of a set of corpus building, management, query and analysis tools. As it is specifically designed for working with spoken data, it is used for various linguistic analyses such as conversation and discourse analysis, language acquisition and multilingualism, phonetics and phonology, dialectology and sociolinguistics. It is most frequently used in discourse and conversation analysis research.

Among its many tools, EXMARaLDA has a transcription and annotation tool (Partitur-Editor), a corpus manager tool (COMA), and a corpus query and analysis tool (EXAKT). These tools were used to construct the CoTY and carry out the analyses. Each tool will be introduced in the following sections.

3.5.1 Partitur-Editor

The transcription and the annotation of the audio data is carried out in Partitur-Editor tool of EXMARaLDA. It provides a 'musical score' interface to work with the spoken data as illustrated in Figure 2 below.

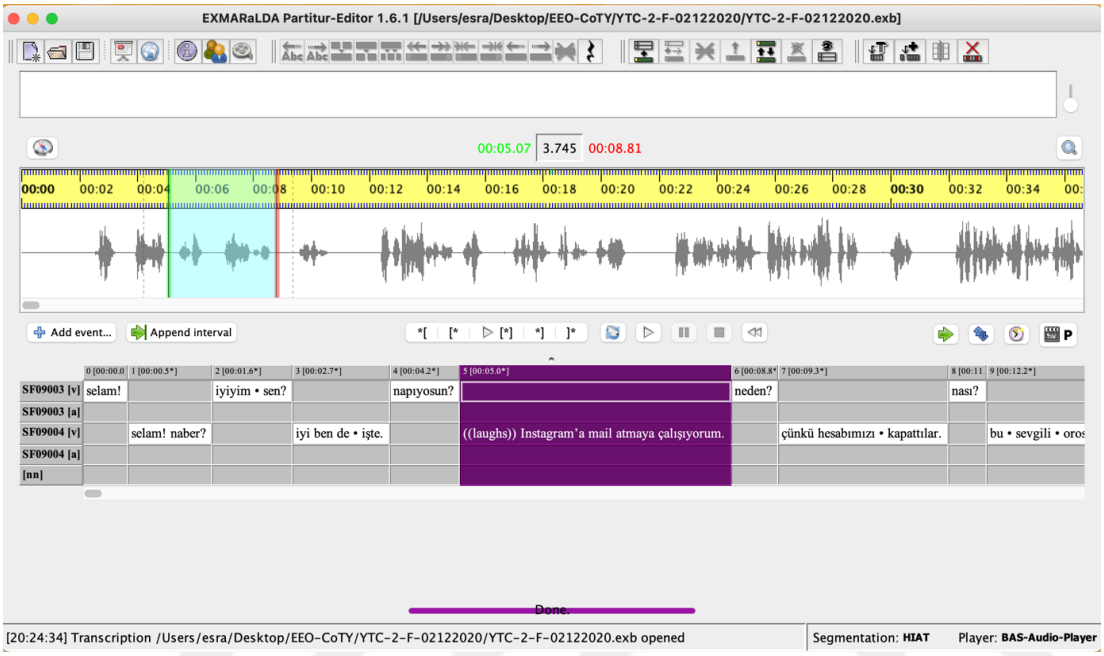


Figure 2 Screenshot of Partitur-Editor interface

Besides the separate tiers (lines) assigned for each speaker in the conversation, additional tiers for annotating paralinguistic features and researcher comments can also be added. Such a multiple tier layout allows the researcher to annotate the overlapping interactional events with clarity and precision.

Partitur-Editor also enables researchers to align the transcription with the audio by assigning time clips to the segments in the transcription.

3.5.2 COMA

COMA is the corpus manager tool of EXMARaLDA. It acts as a bridge between Partitur-Editor and EXAKT through compiling all of the transcriptions and connecting them with the metadata for the queries to be conducted. The total number of speakers in the

corpus, and the overview of metadata of the recordings are presented within the interface of COMA as illustrated below in Figure 3.

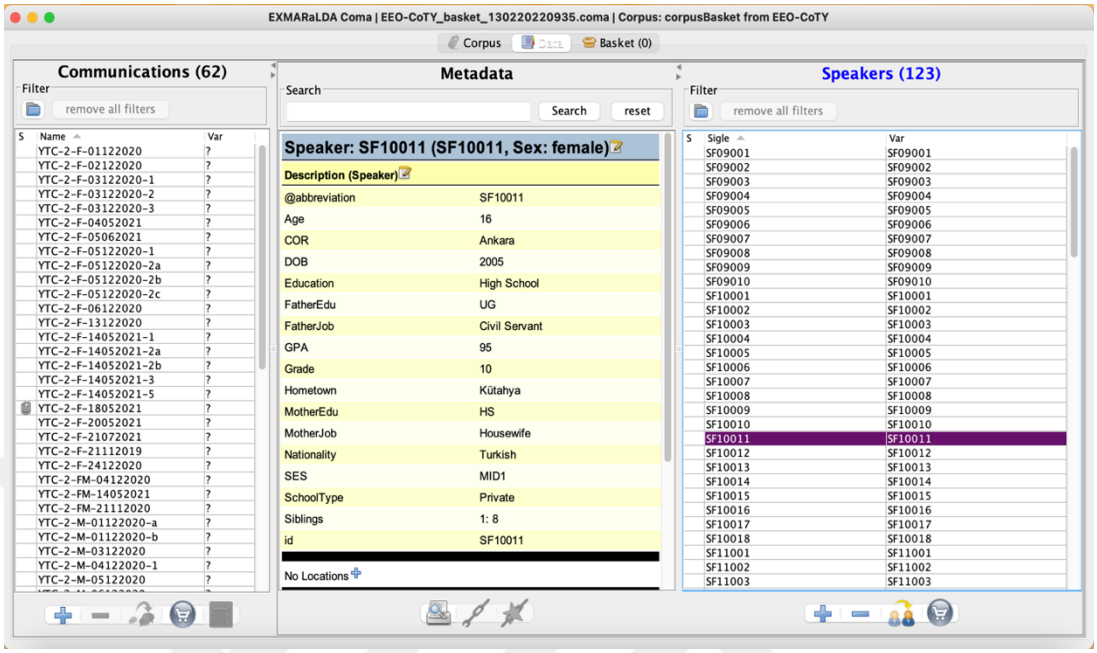


Figure 3 Screenshot of COMA interface

COMA mainly manages two types of metadata, namely communication metadata and the speaker metadata, both of which consist of parameters within the scope of research questions addressed in this study. Please refer to section 3.6 of this chapter for a detailed information about the type and scope of metadata compiled for the CoTY.

3.5.3 EXAKT

EXAKT is the data corpus analysis and concordance tool of EXMARaLDA. It allows for advanced queries via regular expressions, and the query hits are listed as concordance lines. The query results can be automatically or manually sorted depending on the focus of investigation. The expanded contexts and corresponding audio clips for each query hit can also be viewed/played. EXAKT interface is illustrated below in Figure 4.

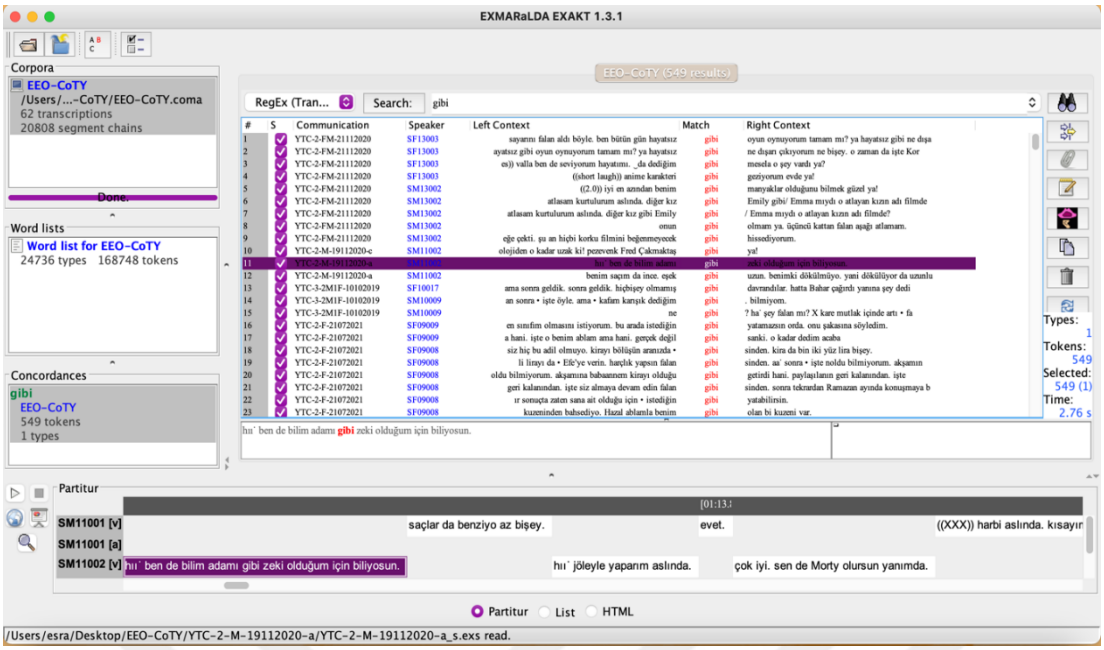


Figure 4 Screenshot of EXAKT interface

One of the advantages of EXAKT tool is that it can correlate the query results with any of the speaker or communication metadata selected. Additionally, the tool provides the researcher with columns for analysis where notes and codes can be written during the analysis stage. Query results can be exported in .xlsx format.

3.5.4 Workflow

After cleaning and categorizing spoken data and metadata, the data was imported to the corpus construction software EXMARaLDA. The corpus construction was carried out by means of the aforementioned Partitur-Editor, COMA, and EXAKT tools of the software. The process was iterative for transcription and annotation stages in Partitur-Editor. Researcher and another transcriber worked as a team to iteratively control unclear utterances, verify speaker identification, and ensure convention consistency. After the final controls were completed for the transcription, annotation, and time-alignment for audio; the data was segmented and incorporated into COMA tool.

Following the stage of processing transcriptions and metadata in COMA, the corpus was constructed and ready to carry out queries and analyses in EXAKT tool. Figure 5 illustrates the procedure followed in the construction of corpus.

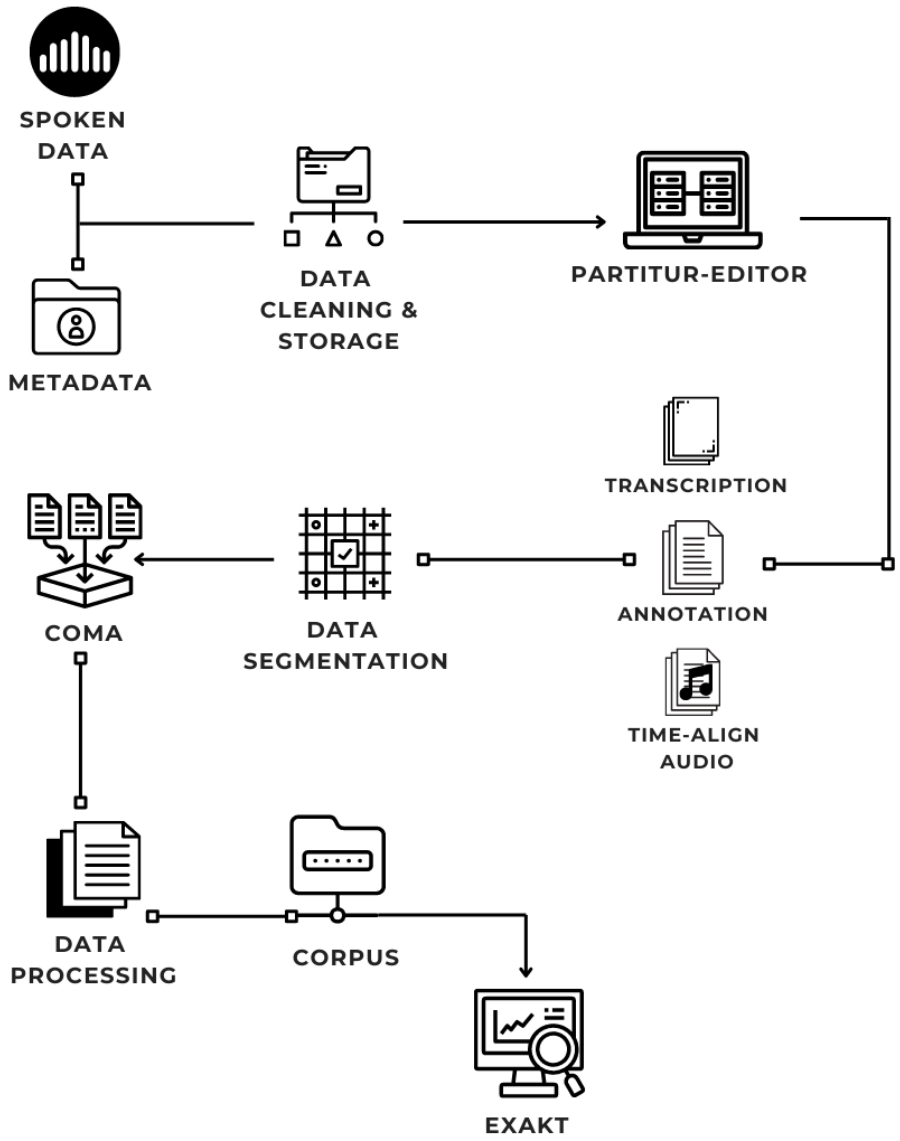


Figure 5 Corpus construction workflow

The detailed information regarding the metadata, transcription and annotation will be presented in the following sections.

3.6 Scope of Metadata in the CoTY

In this study, each audio recording was assigned a unique ID in the corpus and all of its accompanying documents (e.g., transcription file, metadata notes) use the same ID. Similarly, each speaker was assigned a unique speaker ID which is consistent within and

across the recordings in the corpus. ID assignment is important for the purposes of the anonymization and the reliability of the analyses.

This study makes use of rich metadata to contribute to the thick description of the data and the interpretation of the results. The existing corpora of youth language show variety in terms of metadata including age, sex, socio-economic status, ethnicity, first language and other languages known, education level, city of residence, occupation, accent/dialect, relationship to recorder, topic, conversational purpose. The larger the scope of metadata is, the deeper and more detailed interpretations a corpus can offer to the researcher. Therefore, this corpus study utilized rich and diverse components of metadata in its design. That being said, it is also important to note that the components of metadata directly reflect the specific purpose(s) of a corpus. Detailed overview of metadata of existing spoken youth language corpora is presented in Table 5 below along with the comparison of metadata in the CoTY.

Table 5 Comparison of metadata in existing spoken youth language corpora

Metadata	***COLT***	***COLA***	***CORMA***	***Ph@ttSessionz***	***KiDKo***	***JuBe***	***CoTY***
Age	✓	✓	✓	✓	✓	✓	✓
City of Residence	✓			✓			✓
Conversational Purpose							✓
Dialect				✓			
Education Level			✓			✓	✓
Ethnicity	✓					✓	
Father's Education Level							✓
Father's Occupation							✓
Language: first					✓		✓
Language: other spoken/known					✓		✓
Mother's Education Level							✓
Mother's Occupation							✓
Medium of Interaction							✓
Ongoing Activities							✓
Occupation						✓	
Relationship: between speakers							✓

Table 5 (cont'd)

Metadata	*COLT*	*COLA*	*CORMA*	*Ph@ttSessionz*	*KiDKo*	*JuBe*	*CoTY*
Relationship: frequency of communication							✓
Relationship: to the recorder			✓				✓
School: Grade		✓					✓
School: GPA							✓
School: Type							✓
Setting	✓						✓
Sex	✓	✓	✓	✓	✓	✓	✓
Siblings: Ages							✓
Siblings: Number							✓
Socio-Economic Status	✓	✓	✓				✓
Topics	✓		✓				✓

Like its predecessor the Spoken Turkish Corpus, this project prioritised collecting metadata systematically and directly from the speakers in the conversations. In this way, it was ensured that the socio-pragmatic discussions to be integrated into corpus analyses were not decontextualized. Additionally, the contact details of the informants were stored in case the researcher had to confirm anything regarding the metadata even after the data collection phase was completed. As previously explained in section 3.3.3, a single online form was used in which all the metadata items were compulsory to fill in by the participants.

This study made use of two types of metadata: communication metadata and speaker metadata. During the data collection, one of the participants acted as the informant in each conversation, and that individual provided the metadata requested via the Recording Log. The following sub-sections will offer detailed information regarding both groups of metadata used in the construction of the corpus.

3.6.1 Communication Metadata

Communication metadata includes information about the following features of the recording:

- Transcription ID
- Domain*

- Duration of the recording
- Date and time of the recording
- Country and city
- Setting
- Interaction type (Online or Face-to-face)
- Relationship between the speakers
- Frequency of communication between the speakers
- Ongoing activities
- Main topic(s)*
- Additional comments by informant and/or researcher

Among these categories; domain, main topic(s) of the conversation, and ongoing activities were assigned retrospectively (indicated by * above) by the researcher while the rest were collected via Recording Log filled in by the informants. The domain is determined as 'conversation among friends' as a default focus domain for this study. The main topics refer to the conversational topics which were more frequently mentioned or widely spoken between the speakers than any other subject mentioned in a particular recording. Additionally, as a separate analysis, each recording was also coded for all the conversation topics mentioned. Ongoing activities describe the context, in other words, the activities the speakers simultaneously do while talking to each other. The examples would include; online shopping, cooking, riding bicycle, studying, etc. Though the relationship between the speakers is controlled during data collection and set to 'friends', details regarding their relationship such as the duration of speakers' acquaintance, the form of their friendship (e.g., best friends, neighbourhood friends, classmates, former class mates, childhood friends) were also recorded and coded as metadata.

3.6.2 Speaker Metadata

The metadata categories collected for each speaker are as follows:

- Name, Surname (anonymized to Speaker ID while transferring to the corpus)
- Sex
- Date of birth
- Nationality
- Languages used

- First language(s)
- Other language(s) known
- Country and city of residence
- Hometown
- Education level and grade level
- School type
- GPA
- Mother's education level and occupation
- Father's education level and occupation
- Number of siblings and their ages
- Socio-economic status*

Within this group of metadata, only the category of socio-economic status was assigned retrospectively (indicated by * above) by the researcher, the rest of the categories were drawn from Recording Log filled in by the informants. In order to determine the socio-economic status, parents' education level and occupations were used as a source of information. Based on the existing literature and scales for evaluating socio-economic status of various groups of people in Turkey (Kalaycıoğlu, 2010; TÜAD, 2012; Tüzün 2000), a scheme was designed to assign the speakers into six socio-economic levels, namely HIGH-1, HIGH-2, MIDDLE-1, MIDDLE-2, LOW-1, LOW-2 which were collapsed into three main levels: HIGH, MIDDLE, and LOW (See Chapter Four for the details).

3.7 Transcription and Annotation

After the recordings were formatted and assigned their IDs, the very first step was to transcribe them orthographically in order to establish the symbolic representation of the spoken data. Orthographic transcription is vital in the sense that it generates a base transcription for all transcription conventions and further annotations in the corpus. After the base transcription was completed, a second round of transcription in accordance with transcription conventions and annotation scheme was carried out based on the transcription protocol (see Table 6 for the outline of conventions adopted and Appendix F for a comprehensive account with examples). As mentioned previously, the researcher transcribed the pilot data to formulate a detailed and clear set of principles for the transcription protocol to be used. The transcription protocol aims to establish consistency, transparency, and accuracy of the transcriptions (Goedertier et al.,

2000). In order to ensure the consistency, the recent editions of spelling dictionaries published by the Turkish Language Association (Tur. *Türk Dil Kurumu*) and the Turkish Language Organization (Tur. *Dil Derneği*) were consulted for the general rules for spelling of contemporary Turkish language. Though there exist official guides to consult, Turkish shows prominently marked variation in the actual pronunciation of a number of morphemes, words, and expressions. The analysis of pilot data indicated that youth make use of style shifting as a pragmatic and discursive strategy to attain various communicative and social goals in interaction. As a result, this study adopts a transcription protocol which consists of a set of exceptions to the general standard rules for spelling in order to present the spoken language as close to its naturalistic form as possible. As the exceptions constitute a limited set and the variations in the morphemes do not affect the spelling of the root words in Turkish, such an adaptation does not pose any difficulty or limitation for any corpus query or linguistic analysis.

The cases regarding the deviations from standard spelling and selected examples are as follows:

- *Inflectional morphemes:* Rather than standard orthography, the variation in the pronunciation of the morphemes are presented in the transcription. These morphemes include future marker –(y)AcAk as in *yapıcam* for *yapacağım* 'I'll do it' and present tense marker -Iyor + agreement marker as in *yapıyom* for *yapıyorum* 'I'm doing it'.

- *Phonetically reduced forms:* For a limited set of lexemes, phonetically reduced forms are presented in the transcription. These lexemes include the reduced form *abi* for *ağabey ' big brother*'; *vidyo* for *video* 'video'; *dakka* for *dakika* 'minute'; *bi* for *bir* '*one/the/a/an*'; *bişi* for *bir şey* '*something*'; *burda* for *burada* 'here' and other similar variations such as *orda-şurda-içerde-dışarda-nerde*; *di mi* for *değil mi* 'isn't it?'; and diminutive suffix *-cIm* for *-cIğIm* used in address terms.

- *Dialectical variations:* When a speaker style shifts and a variation in pronunciation and/or morphonology is performed, the variation is presented in the transcription. The standard orthography is written as an explanation in the comment tier. The examples include: *gı* for *kız* '*girl*', *yapayrum* for *yapıyorum* 'I'm doing', *gidek* for *gidelim* 'let's go'.

All cases of deviations from standard conventions were documented and compiled in a separate coding log along with examples in order to consult during the transcription process and enhance transparency.

The accuracy of orthographic transcriptions was established through two correction rounds. After the base transcription was completed, a second transcriber who was a Turkish native speaker checked whether the transcription was in accordance with the transcription protocol and whether the speech attributed to the speakers were correct. Following the first round of correction, the researcher did the second round of correction, checked the transcription and corrected the errors if there were any. Employing another transcriber also increased the level of full representation of speech as the unclear utterances could be deciphered through increased rounds of checks.

After a transcription file completed the correction phases, the file was imported to EXMARaLDA Partitur-Editor for the annotation phase which followed HIAT conventions (Rehbein et al., 2004). HIAT (Eng. Semi-Interpretative Working Transcriptions) is a transcription system originally developed by Konrad Ehlich and Jochen Rehbein for the notation of the spoken language in 1970s.

Later in early 2000s, with the development of computer-assisted transcription software EXMARAaLDA at the University of Hamburg, HIAT conventions were integrated into EXMARaLDA Partitur-Editor interface which enables corpus linguists to make use of a single operating system to transcribe and annotate the audio files and align the transcriptions with the audio.

In 2010, researchers of Spoken Turkish Corpus Project at Middle East Technical University adapted the system for Turkish spoken language and a keyboard for Turkish supplement for HIAT was developed and integrated into EXMARaLDA Partitur-Editor.

An overview of conventions used are presented in Table 6 (see Appendix F for the example uses of conventions for Turkish and English).

Table 6 Transcription Conventions (HIAT)

Symbol	*Function*	*Explanation (adapted from the STC Transcription Guideline, 2010)*
•	micro pause	Bullet point sign is used to mark pauses shorter than 0.1 second.
((_._))	timed pause	Pauses equal or longer than 0.1 second are measured and written in double parentheses.
/	repair	Forward slash is used when a speaker corrects, changes a word, or restarts an utterance.
.	falling intonation	Full stop is used to mark declarative utterances and utterances with falling intonation.
?	questions	Question mark is used to mark all utterances and backchannels which are functionally interrogative.
!	rising intonation	Exclamation mark is used to mark utterances with exclamatory function, utterances with rising intonation, greetings, vocatives.
...	cut-off/interrupted utterance (self or other-initiated)	Cut-off sign is used to mark incomplete utterances, self or other-initiated interruptions.
∪	latching	Ligature sign is used when there is not an audible pause between two utterances
-	multi-syllable non-lexicalised or semi-lexicalized units	Hyphen is used for multi-syllable non-lexicalised interjections and other types of semi-lexicalized units such as aggrement markers.
.	non-lexicalised units and paralinguistic features	Superscript dot is sued for non-lexicalized backchannels.
((...))	non-linguistic features	Paralinguistic and prosodic features are marked between double parentheses. Audible actions and background noises are presented between double parentheses.
(text)	uncertain parts	Unclear parts in an utterance are indicated within single parentheses.
((XXX))	unintelligible/inaudible parts	Unintelligible or inaudible parts in an utterance are indicated with three capital X letters put in double parentheses.
<text>	overlaps (markup only in txt file)	Boundaries of overlaps are presented using < >

Annotation had two foci: speech management (pauses and silences, false starts and corrections, overlaps, utterance boundaries) which were marked via HIAT conventions presented above and non-lexical features (paralinguistic features and speech quality) which followed ODT-STD-HIAT (Ruhi, Hatipoğlu, Işık-Güler & Eröz-Tuğa, 2010) conventions. Annotation scheme of ODT-STD-HIAT was adapted in accordance with the particular needs of this study. The scheme was expanded with additional paralinguistic and prosodic features observed in the data. Following Table 7 is the overview of

annotation scheme used and the features generated for this study are indicated by asterisk (*) below.

Table 7 Paralinguistic and prosodic features annotated in the corpus

Paralinguistic features		*Prosodic features*
Assigned to speaker(s)	*No assigned speaker*	
((laughs))	((silence))	((mimicking))*
((short laugh))	((recording cuts off))	((hushes))
((giggles))*	((microphone noise))	((shouting))
((chuckles))	((traffic noise))	((fast))
((snorts))	((sound of turning pages))	((emphatically))
((gasps))*	((sound of dropping stuff))	((singing))
((yawns))	((voices in the background))	((change in tone of voice))
((sighs))	((background noise))	((imitating accent))
((exhales))	((background music))	((softly))
((inhales))	((sound of clapping hands))	((stuttering))
((burps))*	((reads the text))	((syllabifying))
((sniffs))	((sound of mouse clicking))	((lengthening))
((sneezes))	((sound of video playing))	((whispering))
((coughs))	((sound of keyboard))	((murmurs))
((murmurs))	((sound of phone ringing))	((pron Tur))
((kisses))	((sound of shooting a photo))	((pron Eng))
((sings)), ((raps))	((talking on the phone))	
((hums a song))	((wind blowing))	
((clears throat))	((bell ringing))	
((whistles))	((sound of prayers))	
((spluttering))	((subway announcement))	
((stuttering))	((cutting sound))	
((eats))	((sound of water))	
((drinks))	((sound of cutlery))	
((imitating crying))*	((sound of computer))	
((imitating laughter))*	((sound of guitar playing))	
((imitating surprise))*	((writing))	
((imitating cough)) *	((reading))	
((imitating slurping)) *	((eating))	
((imitating grunting))*	((drinking))	

In addition to the paralinguistic and prosodic features presented above in Table 7, audible actions and background noises such as clapping hands, voices in the background, sound of video/music/phone playing, traffic noise, etc. are also annotated in the transcription.

It is also important to note a final layer of annotation which emerged as a result of iterative stages of data transcription and retrospective assignment of topics in the data. This additional annotation consisted of assigning speech events to each conversation in the CoTY. The annotated speech events included conflict talk, gossip talk, troubles talk, storytelling, talking gender, and talking politics. The annotation was implemented at the macro level without marking the boundaries of the beginning and end of these speech events. The main purpose of this preliminary annotation was to create a base for generating sub-corpora to be utilized for further studies. Though limited in terms of its scope, this layer of annotation contributed to exploratory analyses with regard to a specific group of interactional markers (i.e., vague expressions) which will be explained in more detail in Chapter Four.

3.8 Method of Analysis

This study combines a corpus-assisted approach with discourse analysis in order to identify the linguistic patterns of language use and subsequently analyse specific pragmatic and discursive practices manifested in the interaction between the Turkish speaking youth. This section of the current chapter will present the approach adopted for using corpora for linguistic research, and explain the corpus analytical methods employed in this study.

3.8.1 Corpus-oriented discourse studies

Studies using corpora and corpus tools to investigate discourse issues have adapted various names so far. The most frequently referenced distinction belongs to Tognini-Bonelli's (2001) binary terminology, namely corpus-based and corpus-driven linguistics. Tognini-Bonelli's account of corpus-based linguistics refers to corpus linguistics as a distinct method to "expound, test, or exemplify theories and descriptions that were formulated before large corpora become available to inform language study" (2001, p. 65-66) while in corpus-driven approach, "theoretical statements are fully consistent with and reflect directly, the evidence provided by the corpus" (2001, p. 84-85). Within this frame, this distinction requires a linguist to take one of the two contrasting stances: corpus-based linguistics would correspond to adhering to corpus-linguistics-as-method position while adopting a corpus-driven linguistics approach refers to utilizing corpus-linguistics-as-theory in research. The application of these terms in real practice is not as

clear as their definition, though. As McEnery and Hardy (2012) discusses, the ultimate distinction between these two stances relies on the degree to which linguistic evidence from a corpus is used by a researcher, thus corpus-based versus corpus-driven distinction is often unhelpful considering that linguistics research, as is corpus linguistics, is a cyclical and emerging process in which the linguists apply, refine, and redefine existing and emerging understandings of language in a continuum. Additionally, in some cases, linguistic research may require additional forms of data and analysis tools apart from what corpus linguistics offers, such as the use of interviews, field notes, etymological and historical research. Partington (2006) refers this as corpus-assisted analysis which rejects the distinction of corpus-driven versus corpus-based approaches in discourse analysis. In his bibliography of literature on studies related to the use of corpora or corpus linguistic techniques in discourse studies, Gabrielatos (2021) adopts the term corpus-oriented discourse studies to cover the studies focusing on and/or discussing how specific discursive and lexico-grammatical features/patterns contribute to the discourse meaning or construction of particular discourses.

Regardless of the term adopted, combining corpus linguistics with discourse analysis offers several advantages. Baker (2006, p. 10-17) summarizes these advantages as (i) reducing researcher bias, (ii) enabling researcher to recognize the patterns and see the whole picture, (iii) providing a diachronic perspective to discourse by uncovering the resistant and changing discourses, (iv) ensuring triangulation by means of using multiple methods of analysis and/or forms of data. These advantages will be presented with regard to the issues of reliability and validity of the study in the section 3.9 in more detail.

3.8.2 Corpus analytical methods

This section introduces and explains the overview of fundamental corpus techniques used for manipulating corpus data. It is important to note that these techniques are exploratory. Though the statistical procedures produce statistically significant results, the interpretation of the results depends on the research questions.

3.8.2.1 Frequency lists

Once a corpus is at a researcher's disposal to use, the analysis starts with data retrieval. Following the corpus queries via keywords or regular expressions[10], the most fundamental analytical step is generating frequency lists and concordance lines. Frequency of occurrence for the linguistic elements can also be used for uncovering variation in a language (Miller, 2020). *Frequency lists* show the number of occurrences of each token in a corpus. The list can be sorted in alphabetical order or in order of frequency. Frequency lists can be exploited in various ways. For example, they can be compared between two corpora after normalisation[11] in order to identify which words are markedly different or similar in their distribution between two types of registers or two points of time in a single register. Using two normalized frequency lists to compare the frequency of each word in a corpus to its frequency in the other corpus yields a *keyness* value. The comparison of keyness shows *positive* key words (words which are unusually frequent in a corpus compared to the other) and *negative* key words (words which are unusually infrequent in a corpus compared to the other). As with many corpus techniques, frequency analysis is part of an intertwined and cyclical process of corpus analysis. A researcher can select a specific word from the generated list and command the corpus software to produce its concordance lines to explore the discourse constructed around it by means of KWIC analysis.

3.8.2.2 KWIC analysis

Following the corpus queries via a single word, a string of words or a regular expression, the corpus software generates a list of results which shows all the occurrences of data which contain the target item in a corpus are displayed line by line horizontally in a list. The list of hits is called *concordance* and the lines are specifically called *concordance lines*. The target item is labelled as the *node word* and it is positioned at the very centre of each of the concordance lines in the results page. On either side of the node word, a pre-determined number of words which are the words preceding the node word and words

[10] Special characters or strings of characters defined in a corpus and used by researcher to formulate queries by setting the criteria for retrieving data in specified patterns in a corpus.

[11] The process of converting the frequencies of tokens in a corpora into a value of per thousand or per million words in order to allow for comparing corpora of different sizes.

following it are presented. Concordances can be sorted based on the goal of investigation and based on the sorting, an alternative display for the context of the node word is displayed. Exploring the concordance lines to investigate the contextual meaning and use of the node word is called *KWIC* (Key Word In Context) analysis. It allows researchers to identify the frequent collocates of the node word and trail the patterns of these co-occurrences.

3.8.2.3 Co-occurence

Distributional patterning in a corpus is a fundamental analysis to track the semantic and functional similarities between linguistic elements in a corpus (Gries & Durrant, 2020). This pattern is called co-occurrence and it can take the form of a lexical co-occurrence, namely *collocation* (Halliday 1966; Sinclair, 1966), which is the co-occurrence of words with other words and lexico-grammatical co-occurrence, namely *colligation*, which is the co-occurrence of words with grammatical constructions. Linguistic elements which occur together in a corpus are called *collocates* and the degree of frequency of the co-occurrence of the collocates is explored. Contiguous sequences of co-occurrences have been referred by different labels, such as *formulaic sequences*, *lexical bundles*, or *n-grams*, all of which can act as register-specific features. These specific types of co-occurrences can serve different functions in discourse, such as stance indicators, discourse organizers, or referential expressions (Biber & Barbieri, 2007, p. 270). Various collocation measures (e.g., Mutual Information, Log-likelihood, Z-score, Cohen's *d*) are available to calculate the strength of association between the collocates yet the appropriate collocation measure should be chosen in accordance with the focus of investigation.

Collocates also inform a researcher about the patterns of meaning of the target words, among those indicators is a word's *semantic prosody*. A semantic prosody is identified by classifying the collocates of a word into semantic classes. If a collocate has positive connotations, it is noted as having positive semantic prosody, if it has negative connotations, it is marked as having negative semantic prosody. Through corpus methods, change in semantic prosody can be tracked across time periods and different registers.

3.8.2.4 Interpreting Discourse

Aforementioned analytical tools are not ends but rather means to interpret the discourse. A corpus can tell a lot about a discourse when exploited wisely. It can show how the organization and the management of discourse is handled within a register based on the specific parameters (re)sorted by means of the available metadata in a corpus. Among many others, topic management, turn-takings, discourse markers, hedging devices can be identified, classified, and interpreted by means of corpus data and tools. Using corpus tools for discourse analysis yields results that can be both descriptive and explanatory.

In order to achieve a detailed interpretation, it is highly vital that a corpus includes expanded, or ideally full, texts of the discourse. It is important to underline that as the discourse changes over time, the generalizability and validity of the interpretations are always context-dependant because a corpus provides only a snapshot of a phenomenon in question if it is not a monitor corpus.

3.9 Reliability and Validity

As mentioned separately in the previous sections, several measures were taken in order to establish reliability and validity of the study. *Validity* encapsulates the strategies employed to increase the credibility of the research (Creswell & Miller, 2000, p. 125). To ensure this, firstly, a pilot study was conducted in order to test the procedure and data collection tools. Secondly, thick description of the interaction was obtained though detailed metadata. Thirdly, member checking was employed for the interpretation of unclear utterances in the data. Lastly, another transcriber acted as an independent rater during the transcription process of the data. In terms of the corpus construction process, it is important to underline that the sampling frame is a prominent factor which directly affects the validity of the findings a corpus yields.

The design of the sampling frame is intertwined with the issues of authenticity, representativeness and size of a corpus. In this study, each issue was carefully handled and the frame was designed to obtain high levels of authenticity and a maximally representative sample. Integrating contributory public participation model (Shirk, et al.,

2012) into data collection stage was another strategy to increase the level of data precision and accuracy.

As for *reliability*, which is defined as the consistency of analyses and results (Creswell, 2012), a transcription and annotation protocol was established to ensure the consistency, transparency and accuracy of transcriptions. Also, two rounds of transcription checks by two transcribers (researcher and an independent transcriber) were carried out. The speakers were assigned unique IDs to ensure consistency within and across recordings. The literature indicates that ensuring minimally required sample size in a corpus increases stability of analyses regarding identified linguistic tendencies in the corpus. This corpus addresses the concerns of minimally required sample size recommended by Biber (1990). Additionally, using already established transcription conventions (i.e., HIAT), employing a corpus construction and query software (i.e., EXMARaLDA) and using corpus analytical tools (i.e., frequency of occurrence) are the measures which increase the reliability of the queries conducted and discursive patterns identified.

3.10 Ethical Considerations

For this study, the approval from Human Subjects Ethics Committee of Middle East Technical University was granted with the protocol number 150-ODTÜ-2019 (Appendix G). Informed consent forms were prepared for both the participants and the parents/guardians of the participants. Through these forms, the goal, the scope and the procedure of the study were also communicated. All parties were informed that the participation is on voluntary basis and that the study does not involve any items or procedure that might cause any kind of discomfort for the participants. The participants were briefed that they were in no obligation to complete the study and they could leave the study at any point they want. They were also ensured that their names and any personal information would be kept confidential and all of the private data mentioned within the data would be anonymized. The participants were informed about the output of the study (i.e., the corpus) and that the dissemination of the study results would only be used for research purposes. For semi-structured interviews, the meetings were conducted in places and at times convenient for both parties, an environment of physical and psychological comfort were established before the interviews started.

In order to ensure the anonymity of the participants, each speaker in the corpus was given a unique ID, in other words, a pseudonym. The analysis and the reporting of the results made use of these pseudonyms to ensure anonymity.

In this methodology chapter, the design and the stages of corpus construction process were presented in detail. In the following chapter, the findings will be presented.

CHAPTER 4

FINDINGS AND DISCUSSION

4.0 Presentation

This chapter consists of three parts. In the first part, the focus will be on the structural overview of the compiled corpus, The Corpus of Turkish Youth Language (CoTY). Within this first part, the distribution of metadata, the number of tokens and their distribution, the profile of speakers, the forms of interaction between the speakers, settings of communications within the corpus will be presented. Following the structural overview, the second part of this chapter presents the dominant topics and lexical characteristics of the data in the corpus. Finally, in the third part, the most salient linguistic features of spoken Turkish youth language in the CoTY are illustrated under the overarching label of interactional markers categorized into (i) response tokens, (ii) vocatives, (iii) vague expressions, and (iv) intensifiers. In each of these sub-sections of interactional markers, the identified tokens, their distributions, observed patterns, and the ways pragmatic and discursive functions of these linguistic devices are intertwined with the ongoing interaction will be explored and discussed.

4.1 Corpus structure

In this section, the structure of the CoTY in terms of distribution of tokens, demographics of speakers, types and characteristics of interactions are presented. The issues regarding the representativeness of the corpus are also explained. Finally, existing spoken corpora of youth language will be presented in comparison with the CoTY with regard to the scope.

4.1.1 Scope

The current version of the CoTY comprises 168,748 tokens of 24,736 word types[12] within the single domain of informal conversation exclusively among friends. The corpus has 123 unique speakers (62 females and 61 males) and consists of 49 conversations which correspond to 26 hours 11 minutes of dyadic and multi-party spoken interaction. The shortest recording of a conversation is 10 minutes while the longest is 63 minutes.

The language spoken is dominantly Turkish but the speakers integrate words and utterances from English, as well as some from French, Russian and Japanese into their speech. There are 560 tokens (*RF*=3318.55) in English used by 74 unique speakers (38 males and 36 females) in the corpus. A total of 10 tokens occur in the other identified languages.

4.1.2 Speakers

This corpus was designed to consist of speakers whose common denominator is age. With this purpose, corpus data was collected from high school students and recent graduates who were not enrolled in university at the time of the recording. In the Turkish educational context, high schools have four grade levels which are referred to as 9th, 10th, 11th, and 12th grade which respectively correspond to freshman, sophomore, junior, and senior year of high school. Within these grade levels, the age range shows variety, therefore the distribution of speakers with regard to both their grade levels and the ages at the time of recording are coded as separate metadata in the corpus.

Demographics regarding grade levels and corresponding age range of speakers are presented in Table 8 and distribution of tokens is presented in Figure 6 in order to illustrate the architecture of the corpus with regard to speakers in more detail.

[12] In the CoTY, a **token** is defined as the single occurrence of a word. Tokenization is carried out by EXMARaLDA which excludes spaces and punctuation from token count. The methodological constraints currently do not allow for lemmatization of tokens, thus different inflectional forms of a word are counted as separate tokens. A **type** is defined as a unique word in the corpus.

Table 8 Distribution of grade levels and age ranges tabulated by speaker sex

Grade Level	*Age Range*	*Sex*	
		Female	*Male*
9th Grade	**14-16**	10	11
10th Grade	**15-17**	18	17
11th Grade	**16-18**	13	12
12th Grade	**17-18**	13	14
Graduate	**17-18**	3	2
Unindentified	N/A	5	5
Total		*62*	*61*

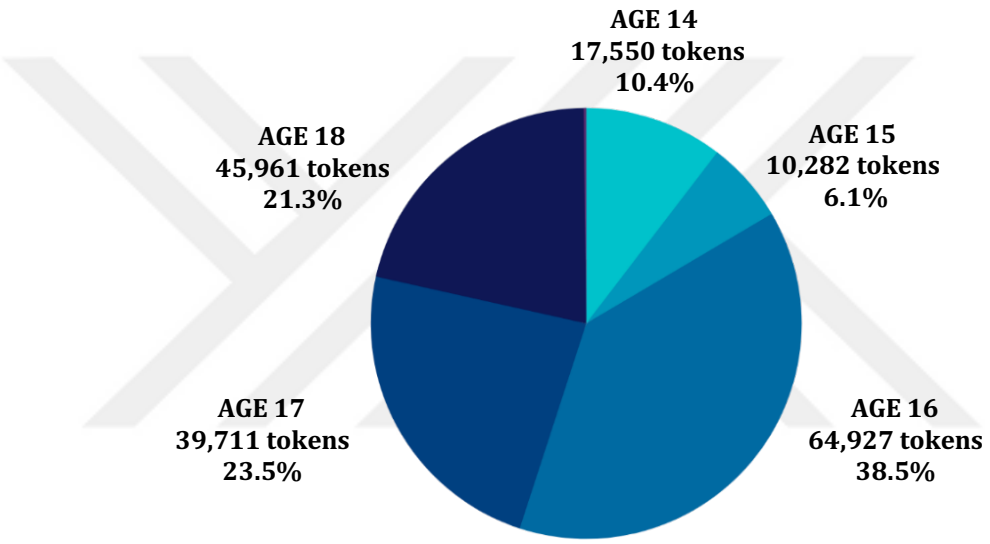


Figure 6 Distribution of speaker ages at the time of recording tabulated by tokens

As Figure 6 shows, the ages of speakers vary between 14 to 18 in the CoTY. 16-year-old speakers constitute 38.5% of all speakers in the corpus with 64,927 tokens (running words) of speech data. The unbalanced ratio between the number of speakers in terms of their age is the outcome of the sampling procedure in which the participants were recruited based on their grade levels as the grade levels have overlapping age ranges. As Table 8 shows, 16-year-old speakers spread over 9th to 11th grades. The saturation of participant numbers in each grade level was determined based on the balance between the number of speakers in each grade which resulted in the highest number of participants and produced tokens in this particular age group.

Though the number of speakers are balanced with regard to sex, female speech make up 58% of the corpus, while 42% of the data is male speech. With regard to sex of the speakers, the CoTY has three types of interactant groups: groups with all-male speakers, groups with all-female speakers, and mixed groups which is made up of male and female speakers. The distribution of data with regard to these groups shows that almost half of the data is exclusively female speech which corresponds to 84,076 tokens (49.8%) in the corpus. All-male speech makes up 26% of the corpus and the remaining 24.2% of the tokens are produced in groups where female and male speakers engage in spoken interaction together (See Figure 7 below).

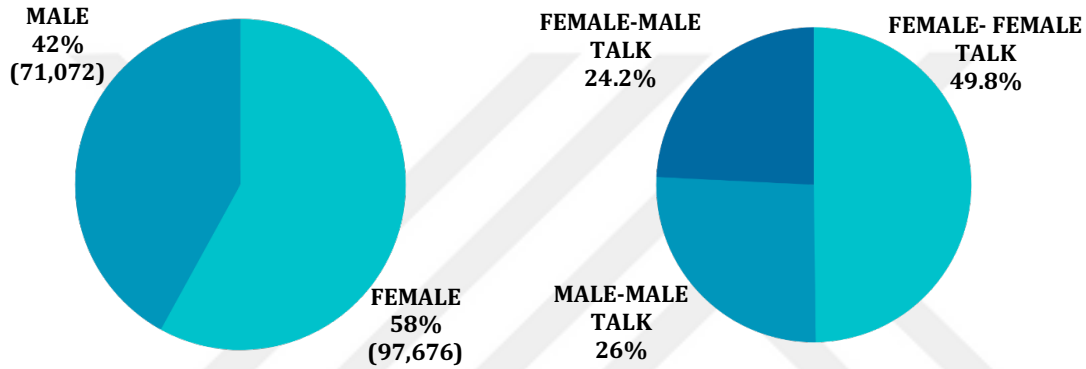


Figure 7 Distribution of tokens by sex of speakers and types of speaker groups

As Figure 7 illustrates, corpus data comprises of both cross and same sex interactions. In terms of the number of speakers in these conversations, each conversation has two or three speakers excluding the speakers who are temporary interactants[13] making unplanned and brief appearances within the course of interaction. Below in Table 9, the detailed structure of conversations in the CoTY with respect to sex of speakers, the number of speakers in each group, the corresponding total hours of speech, and total number of tokens are presented in more detail.

[13] These refer to speakers who made temporary appearances in the recordings such as a friend encountered in the street, a person entering the room, a service provider whose personal data are unknown but who briefly converse with one of the speakers in the recording. In the CoTY, the only information coded regarding these speakers are their sex and the language they spoke. There are 10 unidentified speakers in total and the longest contribution from an unidentified speaker is 72 words and the shortest contribution is 6 words. The total number of words spoken by these speakers makes up 0.2% of the overall corpus.

Table 9 Structure of types of conversations in the CoTY

Type of conversation by sex of speakers		*No. of recordings*	*No. of speakers*	*Hours of speech*	*No. of tokens*	*% of corpus*
same-sex talk	**female**	23	55	12 hr 24 min	84,076	49.8
	male	26	45	7 hr 42 min	43,849	26
cross-sex talk		10	23	6 hr 5 min	40,823	24.2
Total		*49*	*123*	*26 hr 11 min*	*168,748*	*100*

The Ministry of National Education (MoNE) of Turkey conducts educational activities in four levels of education: pre-school, primary school (grades 1-4), secondary school (grades 5-8), high school (grades 9-12), and higher education. High schools in Turkey show variety in terms of their programs. The types of school which have the highest number of students are Science High Schools, Anatolian High Schools, Social Sciences High Schools, Vocational and Technical High Schools, and Religious High Schools[14]. According to 2021 official statistics[15] reported by the Ministry, the majority of high schoolers are enrolled in Anatolian High Schools (92% of the total number of high schoolers in the country) followed by those in Science High Schools (0.6%). In these terms, the distribution of students by high school types in the corpus is representative of that of in Turkey. In the CoTY, 54% of the participants are enrolled in Anatolian High Schools, followed by Science High Schools with 21%. The remaining speakers show variety; they are enrolled in Vocational and Technical High Schools, International High Schools and Social Sciences High Schools. In this sense, data sample of the CoTY reflects a composition of high school types in Turkey except for Religious High Schools[16].

This study also collected demographic information regarding the provinces speakers currently live in and their hometowns[17]. As illustrated in Figure 8 and 9 below, the CoTY offers a wide coverage in terms of both the hometowns and the cities of residence.

[14] https://istatistik.meb.gov.tr/OzetlerKurumTuru/Index (Accessed on June 2022)

[15] https://sgb.meb.gov.tr/www/icerik_goruntule.php?KNO=424 (Accessed on June 2022)

[16] Sampling procedure was not designed to purposefully collect data based on school types but rather it was open to all, yet there were no participants from this specific type of high school.

[17] The concept of hometown in Turkish setting refers to the province from which the parents of a speaker migrated in the past for the purposes of labour, education, etc.

Concerning hometowns, there are speakers from every region in Turkey while only the regions of Northeast Anatolia and Southeast Anatolia are excluded with regard to cities of residence. (See Appendix H for the distribution of all participants grouped under provinces and regions of Turkey).

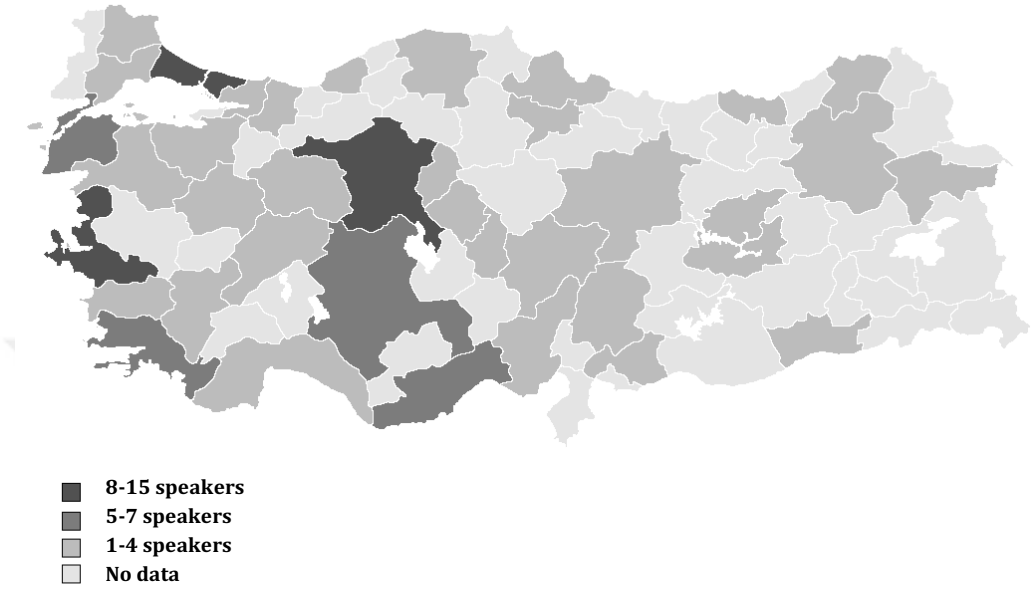


Figure 8 Hometowns of speakers in the CoTY

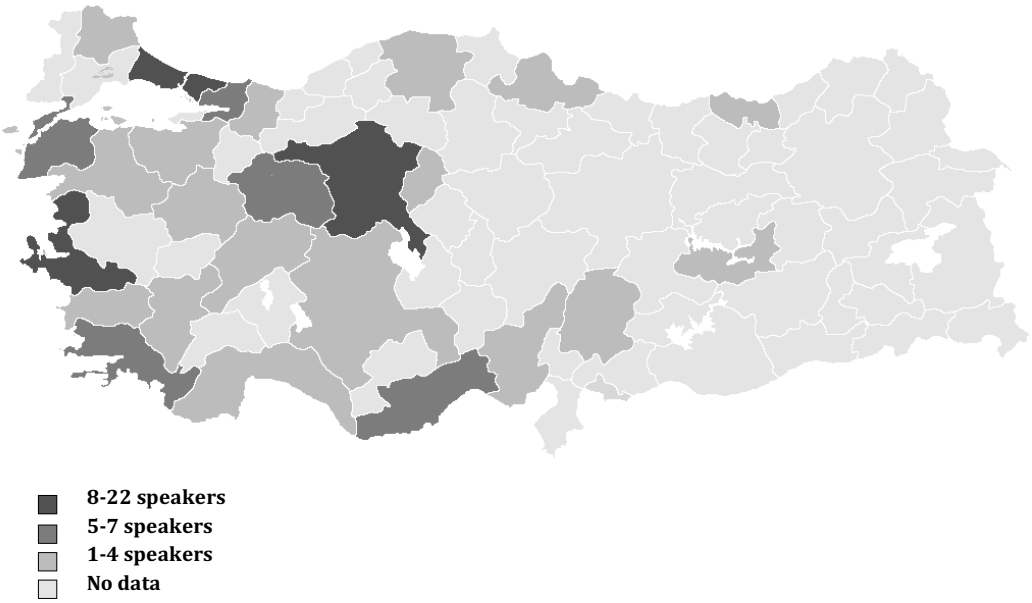


Figure 9 Cities of residence of speakers in the CoTY

In line with 2021 statistics in Turkish Statistical Institute reports[18], the provinces with the highest number of youth population within the age cohort of 15-24 years are Istanbul, Ankara, and Izmir respectively. Similarly in the CoTY, both in terms of cities of residence and the hometowns, the highest number of participants are from Istanbul, Ankara and Izmir.

Though the reported country statistics provide data for a larger group of individuals (15-24 years) than the CoTY (14-18 years), Table 10 shows that the corpus data maximally reflects the distribution of youth according to the most densely populated provinces by youth in Turkey.

Table 10 Youth population in Istanbul, Ankara and Izmir versus the CoTY

	Istanbul		*Ankara*		*Izmir*	
	*population**	*% of total population*	*population*	*% of total population*	*population*	*% of total population*
Turkey	2,263,881	17.5%	826,117	6.4%	573,697	4.4%
CoTY (COR)	19	15.4%	22	17.8%	8	6.5%
CoTY (HT)	15	12%	10	8%	8	6.5%

Population: corresponds to 15-24 years in TÜİK statistics while age range is 14-18 years in the CoTY. COR:city of residence, HT: hometown.

Another demographic layer of metadata in the corpus is speakers' socio-economic background which is a retrospectively coded information. In order to determine the socio-economic status (SES) of participants, a SES scale which was developed for evaluating socio-economic status of Turkish citizens by TÜAD (2012) was used as the main guiding reference for standardizing and coding metadata regarding education levels and occupations of speakers' parents in the CoTY. Based on this categorization, speakers were assigned into three main socio-economic levels which are HIGH, MIDDLE, and LOW and each of these main groups has two sub-levels. This categorization is described and explained in Table 11 below.

[18] https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Genclik-2021-45634 (Accessed on June 2022)

Table 11 Description and distribution of socio-economic status levels in the CoTY

Socio-economic status		*Description*	*No. of speakers*
HIGH	HIGH1	parents hold BA degrees or higher; both have higher managerial/administrative/professional occupations	22
	HIGH2	mother or father holds a BA degree or higher; at least one of them has a higher managerial/administrative/professional occupation	9
MID	MID1	parents have at least high school degrees; both have occupations at public or private sector with steady income	17
	MID2	mother or father has at least a high school degree; one of them have an occupation at public or private sector with steady income	20
LOW	LOW1	parents have a high school or a lower degree; both or one of them work at semi-routine jobs with unsteady income	21
	LOW2	parents have primary school degree; only one of them work at a semi-routine job with unsteady income or both are unemployed	19
Unknown			15
Total			*123*

Levels of HIGH, MIDDLE, and LOW are divided into sub-levels for two reasons. First, socio-economic status of the speakers was determined based on limited number of parameters and the only data source for these parameters are the declaration of the speakers. Secondly, categorizing speakers into only three groups might ignore diversification in terms of their situational factors and enforce rigid interpretations for the cases in which socio-economic status is considered to be a linguistically relevant category.

Though socio-economic status was not among the controlled parameters during data collection, the results depict a balanced distribution of tokens between three main levels which is presented in detail in Figure 10 below.

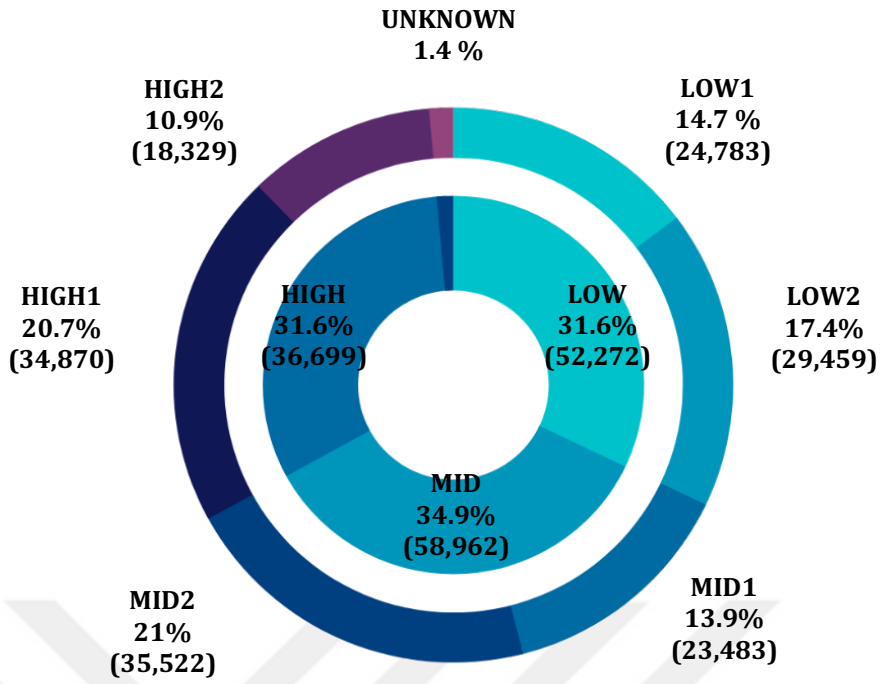


Figure 10 Distribution of tokens with regard to socio-economic status of speakers

The treatment of the category of socio-economic status has been different in existing youth language corpora. Though criticized for inconsistent annotation for socio-economic status (Stenström, 2013, p. 134), the COLT divided speakers into three groups of 'high'; 'middle', and 'low', while the COLAm and the CORMA adopted solely 'high class', and 'low class' as the groups of social class. The COLT made use of the parameters of residential area in London, parents' occupation and whether the parents are employed or not (this data was available only for informants not the speakers) to evaluate the social class while the CORMA reports to be using socioeconomic level of the neighbourhoods the speakers live in as their main source of information for categorizing. Considering six levels sorted into three main groups, the CoTY provides a relatively more fine-grained categorization of socio-economic status without hindering comparability with other corpora.

The results show that the interaction was not confined to conversations between speakers with same socio-economic backgrounds in the CoTY. Rather, each main category interacted with one another.

4.1.3 Domain

As previously presented in Chapter Three, the constructed corpus possesses rich metadata for the profile of speakers and the characteristics of communication among them. The details regarding the communication include the type of interaction, relationship between speakers, frequency of communication between speakers, setting, ongoing activity types, and main topics. The potential pitfall of mainstream corpus-based linguistic investigations is the possibility of decontextualization or lack of contextual cues (Hunston, 2002; Flowerdew, 2008). To address this issue, the CoTY has elaborated metadata regarding contextual details of the interactions in order to enable researcher to employ various levels of analyses in corpus and uncover the facets of the socio-pragmatic meaning within the interaction.

The CoTY has the single domain of 'conversation among friends' which is an informal type of communication taking place in the private domain. Speakers have symmetrical relationships and they exclusively consist of friends with no kinship relations. To obtain an additional facet of the interaction, the speakers were asked to define their relationship to each other under the main category of friends. The speakers constructed their own answers without the limitation of any selective response options, as a result, conceptualization of some speaker relationships are not restricted to a single category. Overall, 40% of the speakers defined their relationship as either best friends or close friends[19], 54% of speakers stated that they knew each other from either class or school. There are also speakers who define their relationships as housemates or neighbours.

Additionally, the speakers were asked to report their frequency of communication (through face-to-face or online communication channels) with each other in their usual daily life. Figure 11 shows the frequency of communication with regard to all groups of speakers and types of speaker groups by sex of the speakers. Regardless of the type of speaker groups, a big portion of speakers (48%) reported to be communicating with each other every day.

[19] In participants' own words, the relationships were depicted as *yakın arkadaş* 'close friend', *en yakın arkadaş* 'best friend', *beşik arkadaşı* 'cradle friends', *doğduğumuzdan beri beraberiz* 'we are together since birth'.

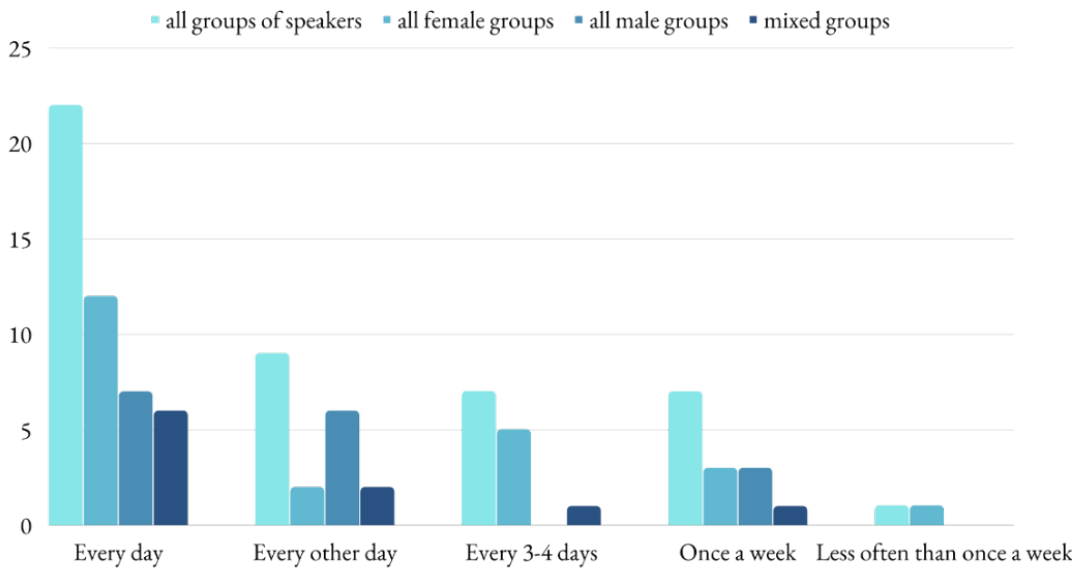


Figure 11 Frequency of communication between speakers within groups

The type of interaction in the CoTY took the forms of face-to-face and online communication. 58% of the interaction was conducted online through online communication channels such as Skype, Zoom, and Discord. Online communication took place in the year 2020 and onwards which was the time COVID-19 pandemic restrictions regarding citizen mobility and education were implemented nation-wide. Face-to-face episodes of communication constitute 42% of the whole corpus and date of the recordings cover a time period from 2019 to 2021. During this period of time, speakers communicated both indoors and outdoors.

Table 12 below presents the distribution of interaction (types and hours of data) corresponding to the pandemic related events (See Appendix C for a detailed timeline of the data collection process and the relevant checkpoints in local and global developments).

Table 12 Types of interaction and duration of recordings tabulated by pandemic related events during data collection period

Year	*Month*	*Events*	*Type of Interaction*	*Duration of recordings*
2019	October-November	▪ Ongoing 2019-2020 Fall Term for Education.	Face to Face (*n*=6)	1 hr 55 min
2020	February	▪ Risk communication campaigns started.	Online (*n*=1)	14 hr 56 min
	March-May	▪ First case of COVID-19 was officially reported on March 10th. ▪ First death due to COVID-19 was officially reported on March 15th. ▪ All schools were closed starting on March 16th. ▪ Distance education started on March 23rd. ▪ Cinemas, restaurants, gyms, concert halls, mosques, malls were closed. ▪ All sports events, scientific and cultural meetings were cancelled. ▪ Government imposed partial curfew for those under the age of 20. ▪ Borders of 31 provinces were shut down except for transiting essential supplies such as food, medical, and sanitary products.	-	
	June-October	▪ A roadmap for normalization period was announced in June. ▪ Curfew restrictions were eased. ▪ Most public spaces including restaurant, swimming pools, and libraries were opened. ▪ University/High School Entrance Exams were held face-to-face.	-	

Table 12 (cont'd)

Year	*Month*	▪ *Events*	*Type of Interaction*	*Duration of recordings*
2020	November	▪ Curfew on people younger than 20 years is reinstated.	Face to Face (*n*=2) & Online (*n*=3)	
	December	▪ The number of daily deaths reached a peak in the country.	Face to Face (*n*=7) & Online (*n*=14)	
2021	January-March	▪ COVID-19 vaccination started. ▪ 8th &12th graders started face-to-face classes at private cram schools to study for high school/university entrance exams. ▪ In-class education started based on the assessment of local risk levels of provinces.	-	9 hr 20 min
	April	▪ Due to the infection rate, nationwide lockdown was implemented by the government.	Online (*n*=1)	
	May	▪ Curfew restrictions for people below 18-year-olds were dropped.	Face to Face (*n*=2) & Online (*n*=10)	
	June	▪ High schoolers returned to school for in-class education with masks on.	Online (*n*=1)	
	July	▪ Curfew restrictions were dropped completely, restaurants resumed activities without restrictions.	Face to Face (*n*=1)	
	August-September	▪ Age for eligibility for vaccination was lowered to 15 years. ▪ All levels of education resumed face-to-face.	-	
	October	▪ The number of infections increased in schools, classes were put in quarantine if a student was diagnosed with COVID-19.	Face to Face (*n*=1)	
Total			***Face to Face (n=19) & Online (n=30)***	***26 hr 11 min***

As mentioned earlier, COVID-19 influenced the type of interaction, as well as the setting and type of speaker activities during communication. The details regarding these parameters are presented in Table 13 below.

Table 13 Type of interaction, setting, and ongoing activities in the CoTY

Type of Interaction	*Setting*	*Ongoing Activity Types*
Online (*n*=30)	**Indoors:** bedroom (*n*=29)	browsing social media and internet, studying, taking photos, playing online games, watching videos, online shopping, singing, drinking, playing guitar, showing books, reading books/emails, sharing screenshots of messages
	Indoor & Outdoor: bedroom & street (*n*=1)	skating
Face to face (*n*=19)	**In doors:** bedroom, kitchen, living room, dorm room, café (*n*=10)	eating, drinking, cooking, organizing, checking bank accounts, playing video games, hanging out in a café
	Outdoors: garden of the house, park, stairwell of the apartment, street, subway (*n*=8)	eating, drinking, cleaning, going to school, walking on the street, smoking, listening to music, watching videos, browsing social media, looking at photos, ordering food online, checking bank accounts, riding bicycle, singing, interacting with animals, playing video games, strolling in the park, solving questions
	Indoor & Outdoor: hairdresser's, street (n=1)	shopping, eating, cleaning

Online communication took place indoors and specifically within personal bedrooms of speakers except for one instance in which two of the speakers were in their personal bedrooms while the third speaker was outdoors, skating in the street and communicating with her friends simultaneously.

As for face-to-face interaction, setting of communication shows more variety. The recordings took place in a variety of indoor locations such as personal bedrooms, dorm rooms, or a café. Outdoor locations included garden of the house, parks, stairwell of an apartment, streets, and subway. Within this type of interaction, there is a single instance of recording in which speakers communicated both indoors and outdoors; they started their conversation in a hairdresser's where they did their internship and later went out to run some errands outside.

While speakers were talking to each other, they were simultaneously engaged in various types of activities as well. They engaged in *activities of daily living* such as eating, drinking, cooking, ordering food, smoking, cleaning, going to school, studying, budget planning; *leisure time activities* such as listening to music, singing, playing guitar, playing video games, watching videos, browsing social media, shopping, skating, riding bicycle; as well as *social activities* such as strolling in the park, hanging out in a café, interacting with animals, taking/looking at photos and playing video games together. As will be presented in more detail in the following section, the variation in activity types is also reflected in the diversity observed for topics in the corpus.

4.2 Topical and lexical characteristics

In this section, topical characteristics of the data will be presented by outlining the main topics and sub-topics coded in CoTY while the lexical characteristics of the corpus will be presented with regard to the results of wordlist comparison and keyness analysis.

4.2.1 Topics

In the CoTY, no directives were given to the speakers with regard to conversation topics, they were simply asked to 'chat as they usually do', thus the topics are non-predetermined and jointly constructed by speakers without any prior planning. The speakers were reminded that all personal data would be anonymized so that they would not have any reservations content-wise and natural flow of topic development was not obtrusive.

No recording in the CoTY starts at the middle of a conversation, therefore identification of topics and contextualization were not problematic. All conversations were retrospectively coded for topics by the researcher. The results are notable in terms of variety of topics in a single conversation (M=12.22, SD=3.53). The data also shows a wide spectrum of topics ranging from daily topics such as news, tv shows, schoolwork to intimate and taboo topics such as romantic relationships, mental health, and issues of sexuality.

So, what do young people prefer to talk about? There are 47 conversational topics identified in the corpus and these topics are thematically clustered under 11 main topics. In terms of main topics, the results show that the most frequently mentioned topics are about entertainment (*n*=151), social and emotional bonds (*n*=133), and education (*n*=111). With regard to sub-topics, speakers present an alternative spectrum of conversational foci, they specifically talk about their friends (*n*=38), social media (*n*=37), COVID-19 (*n*=32), the future (*n*=30) which correspond to the topics that more than half of the groups talked about. The distribution of topics and sub-topics are presented in Table 14 below.

Table 14 Distribution of topics and sub-topics

Topic	*Sub-topic*	*Freq. of occurrences*
Entertainment	Social Media	37
	Show (tv/streaming series, movies, documentaries, reality shows, anime)	33
	Celebrities	28
	Music	15
	Game	14
	Sports	12
	Books	12
	Sub-total	***151***
Social and Emotional Bonds	Friends	38
	Teachers	26
	Family	24
	Boyfriend/Girlfriend	18
	Love Interest	12
	Pets/Animals	10
	Celebrity Crushes/Fanshipping	5
	Sub-total	***133***
Education	Studying/Schoolwork	28
	Teachers	26
	University Entrance Exam	23
	Exams/Grades	18
	Online Education	16
	Sub-total	***111***
Life	Future	30
	Past	21
	Daily Routine	14
	Sub-total	***65***

Table 14 (cont'd)

Topic	*Sub-topic*	*Freq. of occurrences*
Health	COVID-19	32
	Diseases	10
	Doing Sports	5
	Mental Health	4
	Smoking	4
	Stress	3
	Sub-total	***58***
Places	Local Places	21
	Abroad	8
	Weather	5
	Sub-total	***34***
Possessions	Clothes	16
	Electronic Merchandise	13
	Sub-total	***29***
Food/Beverages	Drinks	7
	Local Food	7
	International Cuisine	5
	Desserts	3
	Cooking	3
	Sub-total	***25***
Body Image	Losing Weight/Being Muscular	8
	Height	6
	Hairstyle	4
	Pimples	3
	Getting Tattoos	2
	Shaving	1
	Sub-total	***24***
Politics & News		***17***
Other		***13***
Total		**660**

As mentioned above, the most frequently recurring topics are about forms of entertainment (*n*=151) which covers conversations about traditional and digital media shows and platforms, social media, music, video games, sports, celebrities, books and authors. The results show that entertainment preferences and the content consumed by speakers in the corpus are significantly digital. For example, in terms of shows, speakers prefer shows in online streaming platforms (e.g., Netflix) more than shows on cable television. Additionally, they show big interest in genres of anime and Korean dramas

which are widely disseminated online. As for literature, they follow online writing communities which are built for sharing fan fiction as well as original works (e.g., Wattpad). Celebrities they talk about are not limited to television personalities or movie stars, the speakers frequently talk about influencers (e.g., TikTokers, Instagrammers) in the CoTY as well. Similarly for music, they talk about digital music services (e.g., Spotify) and listen to new forms of audio content such as podcasts. These digital linguistic and semiotic resources are reflected in the linguistic characteristics and discursive practices of youth language which will be presented and discussed in the following sections (see section 4.3).

The second most frequently mentioned topic is social and emotional bonds (*n*=133) and it covers the conversations about friends, teachers, family members, boyfriends and girlfriends, love interests, pets/animals and also celebrity crushes and cases of shipping. The results show that the people they talk about are not confined to the people they are personally acquainted with. The speakers also extensively talk about their celebrity crushes such as actors (e.g., ranging from international stars such as Benedict Cumberbatch, Timothée Chalamet, Zendaya to local figures such as Ezgi Mola, Cem Karaca, Haluk Bilginer) as well as authors (e.g., similarly both Turkish authors such as Oğuz Atay and foreign authors such as J. K. Rowling are mentioned). Additionally, as a concept and form of digital platonic relationship, the speakers exhibit fandom shipping ('*shiplemek*' as linguistically manifested in Turkish data) which is an emotional involvement with the idea that two fictional characters or non-fictional individuals should get involved in a romantic relationship. All of these topics are emotion-laden conversations, speakers construct the discourses of love and affection, frustration and despair, dislike and anger, envy and admiration. As a result, the interactional goals are intertwined with these topics and several communicative acts such as complimenting, disagreement, providing information, convincing manifest within discourse. When it comes to the conversations about speakers' romantic feelings for an individual (be it actual romantic relationships, present or past love interests, celebrity crushes or fandom ships), this topic of conversation is present in 50% of the speaker group conversations in the CoTY.

The third main topic to note is education (*n*=111) which includes schoolwork and grades, their routines for studying, the comprehension problems they face in various subjects of

study, challenges they face within school regulations or national education system as a whole. As the speakers are students in who are not enrolled in tertiary education yet, one of the most frequent topic they talk is national university entrance exams[20], nearly half of the speaker groups (46%) talk about this particular topic in the CoTY. All students who aim to pursue tertiary education are required to take these exams after they complete their secondary education. These exams are multiple-choice tests which are administered nation-wide and take place once a year. As a result, students who are in their junior and senior years prioritize this topic above many others things in their daily routines. Concerning university entrance exams, the speakers talk about their study agendas, their current and target performance, the issues of physical and mental health, the universities they aim to enroll in, expectations regarding college student life, as well as goals and dreams regarding their future occupations. From 2020 onwards, distance education applications and tools became indispensable to all levels of national education in Turkey (refer to previously presented Table 12 for the related timeline). The necessity for swift adaptation to distance education was due to COVID-19 pandemic and as a result, the topic of online education also overlaps with pandemic related issues. One third of the conversations (32%) are related to online education and the speakers usually express their unwillingness to attend online classes and complain about the ineffectiveness of online teaching practices and applications (e.g., EBA platform which is an educational content network developed by Turkish Ministry of National Education).

In addition to the preceding main topics in the corpus, there are also sub-topics which surround the conversations in the CoTY. The most salient conversational sub-topics were identified as *friends* (*n*=38), *social media* (*n*=37), *COVID-19* (*n*=32), and *the future* (*n*=30). These sub-topics are frequently mentioned throughout the analyses which will be presented in the following sub-section of 4.3 in this chapter. Thus in the following part, these sub-topics will be presented in detail along with example excerpts.

As mentioned earlier, the most frequent sub-topic is *friends* (*n*=38) in the CoYT. While talking about or mentioning their friends, speakers do gossip talk or talk about personal

[20] As of 2022, the national university entrance exam in Turkey is called YKS and it consists of three sessions: TYT (*Temel Yeterlilik Testi* 'Core Proficiency Exam'), AYT (*Alan Yeterlilik Testi* 'Field Proficiency Exam'), and YDT (*Yabancı Dil Testi* 'Foreign Language Exam'). All candidates are required who attend TYT which includes questions within the subjects of Turkish, Social Sciences, Mathematics, and Science. Candidates also attend either AYT or YDT based on their field of study and the highher education program they intend to apply. Considering that a total of 3,800,287 students took YKS in 2022, university entrance exams are highly competitive in Turkey.

or shared memories. They sometimes re-enact the story as exemplified in conversation between two female 18-year-old high school graduates from Istanbul as presented in excerpt (1) below[21]. The speakers are reminiscing about their memories in high school, they talk about a number of common friends, and in turn 10, SF13002[22] mimicks one of the people they talk about. As a result, the conversation which included excerpt[23] (1) was assigned the sub-topic of *friends* retrospectively by the researcher.

(1) Y-2-F-13122020

1	SF13001	ee˙ ((name_female1)) şey demişti • ya ben dokuzda da onda da ((name_male))'e güvenmiyodum. çünkü farklıydı ((name_SF13002)). farklı bakıyodu. <sanki böyle… >1> ***err ((name_female1)) said once. well I did not trust ((name_male)) in 9th or 10th grade. because ((name_SF13002)) was a unusual person. he had those weird looks. <as if… >1>***
2	SF13002	<ya kendisi soruyodu. >1> evet zaten ((name_female1)) başından beri söylüyodu yani. ya hatta onuncu sınıfta biz çok güzel bi altılı mıydık? şu an sayamıyorum. bütün sıra vardı ya en önde işte şey/ şeyler oturuyodu. ***<well she was asking. >1> yes ((name_female1)) was telling this from the very beginning. remember we were a very nice six member group in 10th grade? I cannot name them now. there was this whole desk at the very front where those you know who were sitting.***
3	SF13001	evet evet. ***yes yes.***
4	SF13002	hatırlamıyorum. ***I cannot remember (the names).***

[21] Excerpt names refer to IDs assigned to the recordings in the corpus. **Y-2-F-13122020**, for instance, consists of default **Y** letter for 'the youth', the number of speakers (i.e., **2** for two people), sex of speakers (i.e., **F** for female) in the corpus, and the recording date (**13122020**). If there is an additional digit at the end of an ID as in Y-2-M-30112020-**1**, it marks the number of different conversations recorded by different groups of speakers. The letters attached to a final digit such as letter **b** in Y-2-F-14052021-2**b** marks that the recording was stored in parts in the corpus (the recordings which have parts are counted as a single recording in the metadata).

[22] Speaker names correspond to their unique IDs in the corpus. The first letter of all speaker IDs is a default **S** letter standing for 'speaker', the second letter is either an **F** for female speakers or an **M** for male speakers, two-digits following letters refer to the grade level of the speaker (**09** for 9th grade, **10** for 10th grade, **11** for 11th grade, **12** for 12th grade, and **13** for graduates), and the last two digits are ordinal numbers assigned to the speakers in the order of their appearence in the corpus during the data collection phase.

[23] All excerpts used in this dissertation are EXAKT outputs. The transcription follows HIAT conventions (See Appendix F to review the transcription conventions for the ease of reading the excerpts). The personal names are anonymized as ((name_female/male)) in the excerpts. The gloss presented is the idiomatic translation of the data. If there is a lexical item which is the focus of discussion, the related token(s) are presented as underlined in the excerpts.

5 SF13001 ((name_female1)) • ((name_female2)).

6 SF13002 haˑ ((name_female1)) • ((name_female2)) oturuyodu. onların arkasında • senle şey şey ((name_male)) oturuyodunuz.
ah ((name_female1)) and ((name_female2)) were sitting together. behind them, you and ((name_male)) were sitting.

7 SF13001 evet.
yes.

8 SF13002 benle ((name_female3)). falan hani.
I and ((name_female3)). you know.

9 SF13001 evet.
yes.

10 SF13002 böyle arkamız böyle gidiyodu falan. biz çok güzel kaynaşıyoduk. ama ((name_female1)) o noktalarda bile • ((imitating laughter)) ben bi uzak kaliyim ondan. ((imitating laughter)) falan.
and behind us, there were all those people and stuff. we were hanging out really well. but even then ((name_female1)) would say '((imitating laughter)) I prefer to stay away from him ((imitating laughter))' and stuff.

11 SF13001 evet. ve böyle yani ((name_female1))'nın bu çok garip bi özelliği bu.
yes. and this is a very weird characteristic of ((name_female1)).

The second most frequently observed sub-topic is *social media* (*n*=37). Social media platforms mentioned in the corpus are Instagram, TikTok, Slack, Discord, Youtube, Facebook, Tinder, Snapchat, WhatsApp, Twitch. In excerpt below (2) for instance, patterns of social media activity and norms of online behaviour are the main topics within the interaction. In (2) below, speakers who are 16-year-old female high schoolers from Aydın talk about being an influencer and social media etiquette. SF11007 and SF10005 are talking about their followers in social media. SF10005 is telling SF11007 that she noticed that there are some people unfollowing her and even though she feels cross about it, she feels too lazy to dig at it. SF11007 acknowledges this issue and briefly mentions that she experiences the same thing. They both also talk about their high number of followers and pending follower requests on their Instagram accounts. In turn 9, SF11007 states that this issue has effects on her relationship with her school friends, as she often misses the friend requests of people she knew among all the mayhem of messages in her request inbox.

(2) Y-2-F-14052021-1

1 SF11007 ben olsam fark eder miydim diye düşündüm. ben şu an kimseyi kontrol etmiyorum. uygulama falan.

I wonder whether I would have noticed it. I am not monitoring anybody right now. I am not using any applications or anything.

2 SF10005 ben de.
me neither.

3 SF11007 ama fark ederdim herhalde.
but I probably would have noticed it.

4 SF10005 çıkan köpekler var. takipçim azalıyo. görüyorum. ama üşeniyorum • bakmaya. ((short laugh))
there are some bitches who unfollow me. the number of my followers has been decreasing. I am aware of that. but I am too lazy to track it ((short laugh)).

5 SF11007 ben de kanka.
me too, kanka.

6 SF10005 artık. zaten çok fazla takipçim oldu. her ((XXX)) yenilendiğinde Instagram hesapları paylaşıyo falan. birsürü takipçi geliyo. çok fazla oldu. bi elicem bi kaçını.
now. I already have so many followers anyway. every time ((XXX)) is updated, it shares Instagram accounts and stuff. the number of followers increases. there are too many. I'll eliminate some of them.

7 SF11007 kanka bende de birsürü istek var. ee˙
kanka, I have lots of pending follower requests. err˙

8 SF10005 bende de. önüne gelen istek atıyo anam. ‿ben anlamadım ya!
me too. everyone is sending follower requests! I don't get it!

9 SF11007 aynen. şey ee˙ isteklerden şey oluyo. bayağı birikti. altı yüz yedi yüz istek var. o/ bi saniye. o yüzden şey oluyo. tanıdık • mesela okuldan tanıdık biri istek atınca bazen arada kaynıyo. görmüyom ben istek kutusuna sürekli girmeyince. sonra bu da kabul etmedim diye geri şey yapıyo • çekiyo falan. bazı böyle şeyler oluyo. ((short laugh))
exactly. well err because of the requests, that happens you know. it accumulated a lot. there are about six hundred to seven hundred requests. that/ wait a second. that's why the thing happens you know. the acquaintances. sometimes I overlook the friendship requests from people I know from school. I don't notice them when I don't check my inbox regularly. then they withdraw their request thinking that I do not accept it and stuff. things like these happen. ((short laugh))

Another frequently mentioned sub-topic is *COVID-19* (*n*=32) which overlaps with many other topics such as education and daily routines. The speakers talk about restrictions, vaccination, the number of cases, practices of curfew, lockdown, quarantine, the process of normalization, and education with regard to COVID-19. For instance, in excerpt (3) below, 14-year-old male speakers from Antalya are talking about the pandemic measures implemented at their schools. They also speculate about the number of coronavirus cases at their schools. Overall, they are critical towards the inconsistent

implementation of the measures and the official statements regarding the rising number of cases.

(3) Y-2-M-30112020-1

1	SM09003	aynen. yani mesela bizim okulda Koronavirüs tedbirlerinin bazı/ yani her şeyini de dikkat etmiyolar da. yani çoğu mesafeye • hijyene falan dikkat ediyolar. ama mesela derse sınıflara girerken hiç ateş ölçmüyorlar bizde. ***exactly. in my school, they do not completely adhere to the Coronavirus measures. I mean they are careful about the physical distance, the hygiene and stuff. but they never take our temperature while we are entering the classrooms.***
2	SM09004	bizde/ bizde de ölçmüyorlar. sadece ıı˙ ***they don't take our temperature either. only err˙***
3	SM09003	okulun ilk günü ölçtüler. ***they took our temperature on the very first day of school.***
4	SM09004	a/ aynen. bizde yok. her sabah • okula/ okulun içine girmeden önce ölçüyolar. ***exactly. we don't do that. every morning they take our temperature before entering the school.***
5	SM09003	yok bizde hiç yok. bizde ilk gün oldu o kadar. ***no, no such thing at my school. that was only on the first day of school.***
6	SM09004	hmm˙ çıktı mı hiç • Korona vakası? ***have you had any Coronavirus cases at school?***
7	SM09003	oo! hem de ne biçim! bizim bölüm şefi. bizim okuldaki iki öğretmen. bizim sınıftan iki kişi. okulda toplam olarak birsürü vaka çıkmış ama hiç • okulu mokulu/ hiç okulu karantinaya falan da alınmadı. ***ooh! so many! our section chief, two teachers in my school, two students from my class. it is said that there are many cases from the school but they never put the building into quarantine.***
8	SM09004	haa! ***ah!***
9	SM09003	bizim okuldan yirmiden fazla karantina şey/ Koronalı vardır. ‿onu söyliyim ben sana. ***I bet there are more than just twenty people inflected with the virus in my school. let me tell you this.***
10	SM09004	sonra neden • virüsler inmiyo diyorlar. ***and then people wonder why the number of cases does not decrease.***

Conversations about *the future* (*n*=30) ahead include plans about prom night, travelling, plans and dreams of going abroad, vacations or spending time together over the weekend or in summer, the concept of marriage, going to university, future occupations. Speakers talk about the future adopting a positive perspective, often accompanied with

dreams and wishes. For instance in (4), two 15-year-old female speakers from Antalya talk about their dreams. Through successive turns of talk, a vision for the future is jointly constructed in which togetherness is emphasized.

(4) Y-2-F-21072021

1	SF09009	abi bak üniversiteyi kazanırsak ((name_SF09008)) birlikte. İstanbul'da hayatımız çok mükemmel olabilir. bayramdan bayrama! ((chuckles)) ***bro look, if we get to go to the college together ((name_SF09008)), our lives in Istanbul would be perfect. only in holidays (we would visit our family homes)! ((chuckles))***
2	SF09008	ben bayramda bile gelmem. ***I wouldn't return home even in holidays.***
3	SF09009	ya benim öyle sorunlarım yok aslında. ama senin için… ***well, I actually don't have those kinds of issues. but for you…***
4	SF09008	((XXX))
5	SF09009	bak dördümüz. ((name_SF09008)). a· ((name_SF09008)). sensin zaten. ***look, the four of us. ((name_SF09008)). oops you are ((name_SF09008)).***
6	SF09008	((snorts))
7	SF09009	ben. ((name_female1)) ve ((name_female2)) <İstanbul'u kazanırsak. >1> ***I. ((name_female1)) and ((name_female2)). <if we get in to college in İstanbul >1>***
8	SF09008	<abi!>1> hayatımızı yaşarız! gerçekten <hayatımızı yaşarız! >2> ***<bro!>1> we'd live our lives! really <we'd live it!>2>***
9	SF09009	<yemin ediyorum. >2> şuraya yazıyorum. mükemmel olur yaşantımız! abi • yani misafir gelmez. hiçbirimiz misafir sevmiyoruz çünkü. ((chuckles)) ee· gayet yemek de yapabiliriz. ‿bence dördümüzden bi tanesi <yemek yapmayı bulur. >3> ***<I swear. >2> I swear here. our lives would be perfect! bro, no visitors. because none of us likes visitors. ((chuckles)) err we can cook as well. I believe one of us could <figure out how to cook. >3>***
10	SF09008	<ben yaparım. >3> ***<I'll do it. >3>***
11	SF09009	((name_female1)) da yapar. aç da kalmayız. ***((name_female1)) would also do it. we wouldn't starve.***
12	SF09008	senin hiç/ senin hiçbişe yapmicağını hepimiz biliyoruz. ***we all know you won't do any cooking.***

Stenström (2014, p. 10) observed that in the COLT, girls and boys talk about the same topic in different ways. It was also indicated that there are gender exclusive topics such as boys talk about computers and girls talk about their appearance. In the CoTY, on the other hand, all speakers talk about each one of the 47 sub-topics identified in the corpus.

In terms of types of speaker groups, only the topics of pets/animals and electronic merchandise are not present within mixed groups of speakers.

4.2.2 Key concepts and typical vocabulary

To identify what is typical and atypical with regard to spoken Turkish youth language in terms of the key concepts manifested and its typical vocabulary, two complementary investigations were conducted. Firstly, most frequently occurring 100 tokens were generated for the CoTY and the STC and the generated *wordlists* were compared in order to note the preliminary observations regarding the lexis of talk in youth language versus adult language. Secondly, *keyness analysis* which refers to a range of measures and statistics to identify keywords in a corpus. The analysis fundamentally compares the frequencies of words in a target corpus (CoTY) with their frequencies in a reference corpus (STC) and produces a set of words which are typical for the corpus of interest (CoTY). Keywords are useful in the sense that they show "the key concepts in discourses" and "typical vocabulary in a genre/language variety" among other observations (Brezina, 2018, p. 80). For this purpose, keyness analysis was conducted to identify positive keywords in the CoTY.

The Venn diagram in Figure 12 illustrates a comparison between 100 most frequent tokens in general spoken language and spoken youth language of Turkish. The comparison shows that general spoken language and spoken youth language have overlapping tokens (n=79) which comprises of content words, function words as well as non-lexical linguistic items. Additionally, there are corpus-exclusive tokens (n=21 each) for both corpora.



Figure 12 100 most frequently occurring tokens in the CoTY versus the STC

Within the intersection of two corpora, there are overlapping categories of words which can be grouped under **pronouns** (*ben* 'I', *biz* 'we', *sen* 'you', *o* 'he/she/it', *bu* 'this', *şu* 'that', *o* 'that', *orda* 'there'); **determiners** (*bir* 'a(n), one', *bir şey* 'something', *her* 'every', *biraz* 'some', *çok* 'many', *hiç* 'none,no'); **conjunctions and discourse connectives** (*ama* 'but', *çünkü* 'because', *mesela* 'for instance', *ve* 'and', *yani* 'so', *diye* 'as', *d(e)* 'too', *ya* 'if'); **adjectives and adverbials** (*böyle* 'like this', *daha* 'more', *en* 'the most', *güzel* 'beautiful', *hani* 'where', *iyi* 'good' *nasıl* 'how', *niye* 'why', *sonra* 'after', *şimdi* 'now', *zaten* 'anyway'); **postpositions** (*gibi* 'like', *kadar* 'as…as'); **nouns** (*abi* 'elder brother', *adam* 'man', *evet* 'yes', *hayır* 'no', *ne* 'what', *tamam* 'okay', *tane* 'piece', *zaman* 'time'); and **non-lexical particles** (*ee, ha, ya*).

A comparison of corpus-specific lexical items highlights a higher number of adjectives and adverbials in the CoTY list which are *artık* 'anymore', *aslında* 'actually', *aynen* 'exactly', *bayağı* 'extremely', *gerçekten* 'really', *ilk* 'first', *kaç* 'how many', *kötü* 'bad', *sadece* 'only' compared to the STC list which only comprises of the tokens *aynı* 'same', *başka* 'other', and *şöyle* 'that way'. All of these lexical items have discourse organizational functions in Turkish. Another category of lexical items in the list which the CoTY show more variety is nouns which include *kanka* 'dude', *lan* (slang expression generally used as a vocative, a derivative of *oğlan* 'boy'), *oğlum* (literally 'my son', closest English equivalent would be 'man' or 'dude'). The STC list, as opposed to that of CoTY, only includes *hocam* 'my teacher' within this group. These specific lexical items are in general used by speakers to address their interlocutors in spoken Turkish.

To explore the typicality in the CoTY further, *word frequency classes* for the CoTY and the STC are compared by using Log2 calculation (Perkuhn et al., 2012) which works in a similar way to the %DIFF calculation (T. Schmidt, June 2022, personal communication). Frequency class comparison was chosen on the basis of its feasibility and compatibility with EXMARaLDA word lists output. The target corpus was set as the CoTY and the reference corpus used was the STC, top 1000 most frequent words were calculated using word class frequency method. This comparison of frequency classes across corpora yielded a keyness list for the CoTY. Tokens in the list are grouped under two main categories: (i) concepts related to daily life and education, and (ii) function words[24]. Both lists provide observations regarding different aspects of the corpus. Keywords in domain (i), for instance, illustrate the main topics and concepts manifested in interaction among Turkish speaking youth which is presented in Table 15 below (see section 4.2.1 for more details on the identified topics in the corpus).

Table 15 Keywords in the domain of daily life and education

Category	*Keyword**	*Gloss*
Daily Life	**dizi**	*series*
	fotoğraf	*photo*
	Korona	*Coronavirus*
	şarkı	*song*
	sezon	*season (of series)*
	spor	*sports*
	takip	*a follow (social media)*
	vidyo	*video*
Education	**fen**	*Science*
	fizik	*Physics*
	hoca	*teacher*
	İngilizce	*English*
	kimya	*Chemistry*
	konu	*subject*
	matematik	*Mathematics*
	okul	*school*
	online	*online*
	tarih	*History*
	TYT	*(abbrev.) National University Entrance Exam*

*sorted alphabetically

[24] Verbs are excluded from the list as conventions for inflectional forms of verbs transcribed for the CoTY and the STC differ.

Keywords in domain (ii) corroborate the observation made based on comparison of most frequently occurring tokens in the STC and the CoTY (Figure 12) and provides a more elaborated view on the salient tokens which are typical of youth talk in the CoTY. These keywords are grouped under their dominant functions[25] in the corpus and and the list is presented in Table 16 below.

Table 16 Keywords in the domain of interactional markers

Category	*Keyword**	*Gloss*
Intensifier	**aşırı**	*excessively*
	bayağı	*excessively*
	cidden	*seriously*
	gerçekten	*really*
	full	*full*
Response Token	**Allah'ım**	*My God*
	aynen	*exactly*
	oha	*whoa*
	of	*ugh*
	okey	*okay*
	uf	*ouch*
Swear Word	**amına**	*node for vagina-plus swearing expressions*
	anasını	*node for mother-plus swearing expressions*
Vague Expression	**bişey**	*something*
Vocative	**aga**	*derivative of 'elder brother'*
	arkadaşım	*my friend*
	aslanım	*lit. my lion*
	bro	*bro*
	gerizekalı	*idiot*
	kanka	*dude*
	la	*derivative of 'boy'*
	lan	*derivative of 'boy'*
	oğlum	*my son*
	salak	*stupid*

*sorted alphabetically

Adding on the observations regarding the comparison of most frequent words in both corpora, keyness analysis layed the basis for foci of further exploration. The categories

[25] Separate KWIC analyses for first 10 concordance lines of each keyword was conducted to identify the dominant functions.

identified in Table 16 above shaped the groups of linguistic items identified to be investigated in the corpus. These groups will be presented as *interactional markers* in this study and the following section will provide detailed information.

4.3 Interactional markers

Spoken language is highly dynamic in nature and participants in conversation are active agents in the co-construction of the conversation. In this joint endeavour, participants need to maintain the conversation while attending to both discourse and relational management (Rühleman, 2007). As a result, spoken language contains multiplicity of linguistic elements and exhibits a more fragmented structure and as well as high degrees of interactional versatility and vitality. These characteristics call for a relational perspective to examine the spoken discourse. Adopting Roulet's (1980) term 'interactional markers', Ruhi (2013) expands on the notions of discourse/pragmatic markers and offers a comprehensive view of markers which can contain lexical devices (i.e., entities referred interchangeably as pragmatic markers, discourse markers), non-lexical elements (i.e., backchannels, laughter), prosodic features (i.e., change in tone of the voice) and gestures in spoken interaction.

The literature on youth talk has highlighted several features of the language of young speakers and these features can be grouped under the labelling of interactional markers such as pragmatic markers, intensifiers, vocatives, invariant tags, swear words, taboo words, discourse connectives to name a few. Yet for the purposes of this study, the selected groups of linguistic entities were identified based on the keyness analysis conducted in the CoTY[26] in order to present an account of salient interactional markers within the corpus.

Following Ruhi's encompassing classification of *interactional markers*, this study investigates the notable groups of linguistic entities categorized as **(i) response tokens, (ii) vocatives, (iii) vague expressions,** and **(iv) intensifiers** under this classification.

[26] As presented in Tables 15 and 16, keywords were grouped into lexical items within the domains of *daily life and education* and *interactional markers*. The keywords within the domain of interactional markers were clustered into categories of intensifiers, response tokens, swear words, vague expressions, and vocatives. As the category of swear words overlaps with the other categories in the CoTY, they are presented within and across the analyses of intensifiers, response tokens and vocatives in this study.

For each category of interactional markers; types of tokens, their distribution in the corpus and their patterns will be presented and the pragmatic functions of salient items will be discussed along with excerpts from the CoTY. The very first group of interactional markers is *response tokens* which will be presented in the following section.

4.3.1 Response tokens

Spoken discourse is an act of co-construction yet the literature often positions the linguistic analyses with regard to speaker behaviour. Structure and patterns of listenership behaviour, though, suggest that listenership orients more towards affective and relational space in interaction, rather than simply giving acknowledgement (McCarthy, 2002). There is a plethora of labels used for the allegedly 'short' linguistic devices a listener uses with responsive functions in interaction thus the scope and the defining boundaries of these tokens diverges vastly. For this reason, the following section will outline the definition adopted in this study and the scope of tokens determined as the focus of analysis.

4.3.1.1 Defining response tokens

Response tokens are small multimodal signals which are frequently discussed in terms of their communicative functions in spoken interaction. These signals have also been termed as 'backchannels' (Yngve, 1970); 'continuers' (Schegloff, 1983); 'minimal responses' (Fishman, 1983); 'acknowledgment tokens' (Jefferson, 1984); 'reactive tokens' (Clancy et al., 1996); 'response tokens' (Gardner, 1998, 2001; McCarthy 2002); and 'discourse particles' (Aijmer, 2002) among many. Though the definitions overlap and diverge in their scope, existing studies underline that these linguistic devices are multifunctional (Gardner, 2001; McCarthy, 2003; O'Keeffe & Adolphs, 2008) and the most frequently examined functions define them as the marker of understanding/agreement and maintenance of current turn of the speaker.

Focusing on turn-taking architecture, McCarthy (2002) utilizes Sinclair and Coulthard's (1975) labelling system for sequential moves of initiating, responding and follow-up within the structure of conversational exchange to identify response tokens in interaction. McCarthy (2002) focuses on response and follow-up moves and examines

response tokens which constitute the whole turn or those which are in turn-initial positions in more extended responses.

Response tokens are part of active listenership behaviour; thus, they frequently overlap with the speaker's talk (Aijmer & Rühleman, 2015) but do not take over the speaker turn (Duncan and Fiske, 1977; Heinz, 2003; O'Keeffe & Adolphs, 2008; Schiffrin 1987; Tottie, 1991). They can manifest as single-word particles but they can also occupy a longer string of response along with other tokens preceding them or they exist in clusters (McCarthy, 2002). As a result, it is sometimes difficult to differentiate brief utterances from whole turns and determining whether the response tokens have any role in challenging the speakership or not remains controversial (Duncan & Niederehe, 1974). Still, the observation made by Gardner (1998) that they exist "between speaking and listening" points at the active role of response tokens in co-construction of discourse. What constitutes a response token, then, should be clearly defined and justified within the patterns of listenership that is under investigation.

Though their boundaries are not always clear, response tokens are often grouped into minimal and non-minimal tokens (Fellegy, 1995; Fishman, 1978; Gardner, 1997, 2001; Schegloff, 1982; Tottie, 1991). For English, minimal response tokens consist of short utterances (e.g., *okay*) and non-lexical vocalizations (e.g., *mm-hmm*) while non-minimal response tokens include adverbs or adjectives (e.g., *really good*) and short phrases (e.g., *that's excellent*). At this point it is important to state that the present study categorizes response tokens into two groups for Turkish: non-lexical response tokens and lexical response tokens. The details regarding this classification is explained in more detail in section 4.3.1.3.

4.3.1.2 Brief overview of related work on response tokens

The overview of studies on lexical particles which exhibit functions of response tokens in corpus-based spoken discourse will be briefly presented in two sub-sections: recent work on Turkish and youth language research.

4.3.1.2.1 Response tokens in Turkish

Studies on response tokens in Turkish mainly include corpus-based investigations of individual lexical items. Among the works based on the STC data, the comparative approaches to classify and identify the pragmatic functions of a number of tokens stand out. Ruhi (2013) provides a prominent discussion with regard to the fuzzy boundaries of terminology adopted for small linguistic markers in spoken interaction and the need for a comprehensive approach to explore the affective dimension of these tokens within a relational dimension of interaction, namely (im)politeness theories. Ruhi (2013) discusses the interactional functions of *tamam* and *peki* (both literally meaning 'okay' in English) which mark acknowledgement and (dis-)agreement while simultaneously indexing (im)politeness in spoken Turkish. Özcan's (2015) master's thesis combined a conversational analytical perspective with a corpus-based approach to explore the interactional features of lexical token *evet* 'yes' and non-lexical token *hı-hı* in the STC. The analysis identified different patterns in terms of overlaps, sentence positions, domains, and intonational features of these tokens with regard to functions of approval, agreement, continuation, question-response and divergence. Altunay and Aksan (2008) focused on *hayır* 'no' and *yok* (lit. non-existent, negative existential expression) and examined their textual and interactional functions as pragmatic markers in conversation. Bal-Gezegin (2013) compared functions of lexical token *hayır* versus non-lexical token *cık*, and the results show that even though these tokens mainly have similar functions in spoken discourse, they also possess exclusive functions. Both devices have the function of responding to request for information and disapproval/disagreement while *hayır* has the exclusive functions of connective, response to a request/offer/command, metalinguistic negation and *cık* has the exclusive functions of pre-signalling a negative statement.

Apart from aforementioned works focusing on specific tokens, Aytaç-Demirçivi's (2021) doctoral dissertation study stands out as an extensive work on backchannel classification of contemporary spoken Turkish. 150,494-word sub-corpus of the STC was used to identify all the lexical and non-lexical instances of backchannels along with their functions. The study groups backchannels into two main functions of keeping the conversational flow and showing attitudes. Under each main functions, the distribution of each sub function is presented with regard to age group and gender of speakers in interaction. The study highlights the observation that all-female groups use

backchannels more than other speakers do and the functions they most frequently utilize are approval and agreement.

Based on the TNC data, Kaynarpınar (2021) investigated the approval markers under Ruhi's (2013) classification of interactional markers within the framework of (im)politeness and discussed a range of linguistic devices such as *aynen* 'exactly', *doğru* 'right', *elbette* 'sure', *iyi* 'good' which also overlap as instances of response tokens in Turkish.

4.3.1.2.2 Response tokens in youth talk

Even though there is extensive research on various linguistic markers in youth language, those specifically examining response tokens are quite few in number and they display a variationist and corpora-based approach to the investigation of response tokens.

Stenström's (2014) extensive work investigated the linguistic devices under a comprehensive umbrella term 'pragmatic markers' in a cross-linguistic perspective by comparing the talk of London teenagers in the COLT with the Madrid teenager talk from COLAm. Among the pragmatic markers identified, a group of tokens were grouped under directive and reactive moves, with reactive moves corresponding to responding moves in conversational exchange. Among these markers, Spanish *vale*, *no* and English *yeah, okay* and *right* had the same reactive functions across corpora. The results indicated that Spanish equivalent of response utterance *I know* in English youth talk did not have the same function in the COLAm data. Additionally, laughter and interjections are also highlighted as the most common response signals in English and Spanish youth talk. The study also mentions 'reaction signals' identified for Spanish youth language and groups them into the functions of agreeing, objecting and showing surprise.

Investigating the functions of various intensifiers in the language of British adults and teenagers based on the SCoSE, the DCPSE and the COLT, Núñez Pertejo and Palacios Martínez (2014) focused on lexical items *absolutely* and *totally*. Results showed that *absolutely* as an affirmative response token occurs more frequently in adults talk compared to language of teenagers. The study suggested that young speakers of English find *absolutely* too formal to use, as they prefer using *ok, cool, I know, yeah* to indicate agreement in interaction. As for *totally,* corpus analysis indicated that *totally* is used as a

response token which has the function of confirming the immediate statement of the speaker among its other functions. Similarly, Aijmer (2011) mentions that *totally* is an emphatic response token in American English and a "salient feature of teenage talk" (p. 168) which exhibits hyperbolic and boosting functions.

Adolphs and Carter (2013) generated two sub-corpora of young women's speech from the LCIE and the CANCODE in order to conduct variationist research on the use of response tokens in British and Irish English. Both corpora comprised of 10,000-words each and the data was two-party or multi-party talk between close female friends (mostly students) around the age of 20 years. The data was qualitatively analysed within the scope of taxonomy of response token functions developed by the researchers. Even though British English data showed a higher number of response tokens, functional analysis showed similar patterns in both corpora, convergence was the most frequent function followed by the function of engagement.

In order to explore the intersection of response tokens and Turkish youth talk, the following section will present the results of the corpus analyses which illustrated the types, the distribution, the patterns, and provide discussions regarding the salient functions of response tokens used by Turkish speaking youth.

4.3.1.3 Findings: Response tokens in the CoTY

Turkish is an agglutinating language, thus the boundaries for a class of lexical response tokens requires a different categorization than what has been adopted for English so far. To elaborate, expression *I see* which is categorized within the class of non-minimal responses in English corresponds to a single-word token *anlıyorum* 'I understand' in Turkish. Similarly, the results show that single-token responses in the CoTY can include nominals such as *muhtemelen* 'probably' and inflected nominals such as *Allahım* 'my God', verbs inflected for tense/aspect/modality and person such as *biliyorum* 'I know'. There are also short swearing expressions which are not one-word tokens but typically have a node word such as *ana* 'mother' producing one-word or two-word swearing expressions in the data. These lexical response tokens are clustered as mother-plus swear words, and only the node word is included in the lexical response token list. This is why for the purposes of this study on Turkish, rather than adopting the minimal and

non-minimal categorization, response tokens in this study are grouped into lexical and non-lexical groups of tokens.

The first group consists of *non-lexical response tokens* which refer to short vocalizations such as *hı-hı* (closest English equivalent would be *mm-hmm*) and interjections with both established standardized spelling such as *haydi* 'come on' and those which have variants for representation in the literature such as *hah.* As the interjections can point a change in "the state of knowledge, information, orientation and awareness" (Heritage, 1984, p. 299), they also organize everyday talk (Yngve, 1970) in addition to expressing emotions. As a result, this study treated interjections as candidate response tokens and the analyses yielded a number of interjection which were used as non-lexical response in the corpus.

The second group consists of *lexical response tokens*, refer to one-word lexical responses such as *evet* 'yes', repetitions of these lexical responses such as *evet evet* 'yes yes', premodified responses such as *kesinlikle evet* 'absolutely yes', and clusters of lexical response tokens such as *evet aynen* 'yes exactly'.

As for the analysis, the study adapted McCarthy's (2002) corpus-based approach to identify response tokens and examined the pragmatics of these linguistic devices within O'Keeffe and Adolphs's (2008) taxonomy of functions. To identify response tokens, wordlist of most frequently occurring 1000 words for the CoTY was generated using the EXAKT tool and the list was examined manually to mark the lexical items as candidate response tokens based on both the existing literature in Turkish (see 4.3.1.2.1), the list of backchannels identified in the STC design (Ruhi et al., 2010), and the emergent list of tokens identified during the corpus construction stage by the researcher. Each potential response token was then analysed within the KWIC, concordance lines and expanded contexts of tokens were qualitatively investigated. Stand-alone tokens and tokens in turn-initial positions in response and follow-up moves that are not turn-yielding were counted as response tokens. This analysis also yielded the identification of clusters as response tokens. For the purposes of this study, only two-word clusters and reduplications are included in the list.

In the following sections, the results will be outlined in two main categories: lexical and non-lexical response tokens. Following the presentation of distribution of tokens in each

category, due to space limitations, a selection of response tokens will be presented along with excerpts from corpus data. Additionally, swear words, words from religious domain and slang words and expressions which are identified to be used as response tokens will be presented. Finally, a register-specific response token *aynen* 'exactly' will be the focus of in-depth analysis as it is identified as one of the keywords in the CoTY and at the same time the second most frequently occurring lexical response token in the whole corpus.

4.3.1.3.1 Non-lexical response tokens

There are 36 types of 1305 tokens of non-lexical response tokens in the corpus. Table 17 below lists the types of tokens and their frequencies tabulated by speakers in the corpus. In the table, total number of tokens retrieved from the corpus (*TN*), the absolute frequencies of total number of identified response tokens (*AF*) along with the relative frequencies (*RF*) per million is presented in a descending order. For each non-lexical response token, total number of unique speakers using that token, and the distribution of female and male speakers are also presented to illustrate the extent each token is used by Turkish speaking youth in the study.

Table 17 Non-lexical response tokens tabulated by frequencies and speakers

Rank	*Non-lexical response token (type)*	*No. of tokens*				*No. of addressers*					
		All occurences		*Non-lexical response tokens*		*All*		*Female*		*Male*	
		TN	*RF*	*AF*	*RF*	*N*	*%*	*N*	*%*	*N*	*%*
1	hı-hı	375	2222.25	**337**	1997.06	64	52	41	66	23	14
2	hıı	213	1262.24	**147**	871.12	62	50.4	32	52	30	18
3	ha	337	1997.06	**146**	865.20	68	55.3	32	52	36	22
4	hı	176	1042.98	**126**	746.68	52	42.3	32	52	20	12
5	haa	146	865.20	**86**	509.64	33	26.8	14	23	19	12
6	hmm	116	687.42	**59**	349.63	31	25.2	20	32	11	6.7
7	ee	436	2583.73	**42**	248.89	24	19.5	9	15	15	9.2
8	oha*	74	438.52	**40**	237.04	25	20.3	12	19	13	7.9
9	ya	880	1514.88	**40**	237.04	31	25.2	18	29	13	7.9
10	aa	80	474.08	**34**	201.48	24	19.5	12	19	12	7.3
11	be	93	551.12	**30**	177.78	20	16.3	11	18	9	5.5
12	hah	81	480.01	**25**	148.15	26	21.1	16	26	10	6.1
13	ay	136	805.94	**22**	130.37	19	15.4	16	26	3	1.8
14	ah	39	231.11	**21**	124.45	15	12.2	7	11	8	4.9
15	ha(y)di	154	912.60	**19**	112.59	15	12.2	5	8.1	10	6.1
16	cık	84	497.78	**16**	94.82	12	9.76	3	4.8	9	5.5
17	hee	30	177.78	**15**	88.89	11	8.94	4	6.5	7	4.3

Table 17 (cont'd)

Rank	Non-lexical response token (type)	No. of tokens				No. of addressers					
		All occurences		Non-lexical response tokens		All		Female		Male	
		TN	RF	AF	RF	N	%	N	%	N	%
18	oo	31	183.71	**14**	82.96	13	10.6	4	6.5	9	5.5
19	yaa	27	160	**13**	77.04	10	8.13	9	15	1	0.6
20	ı-ıh	22	130.37	**9**	53.33	8	6.5	3	4.8	5	3.1
21	hıh	22	130.37	**8**	47.41	8	6.5	4	6.5	4	2.4
22	vay	44	260.74	**8**	47.41	6	4.88	3	4.8	3	1.8
23	of	159	942.23	**7**	41.48	6	4.88	4	6.5	2	1.2
24	uf	78	462.23	**7**	41.48	4	3.25	2	3.2	2	1.2
25	hayda	11	65.19	**7**	41.48	3	2.44	1	1.6	2	1.2
26	he	22	130.37	**6**	35.56	5	4.07	1	1.6	4	2.4
27	wow	8	47.41	**4**	23.70	3	2.44	3	4.8	0	0
28	eh	3	17.78	**3**	17.78	3	2.44	1	1.6	2	1.2
29	eww	4	23.70	**2**	11.85	2	1.63	0	0	1	0.6
30	heh	12	71.11	**2**	11.85	2	1.63	0	0	2	1.2
31	oho	6	35.56	**2**	11.85	2	1.63	1	1.6	1	0.6
32	öf	25	148.15	**2**	11.85	2	1.63	0	0	2	1.2
33	üf	17	100.74	**2**	11.85	2	1.63	2	3.2	0	0
34	yuh*	8	47.41	**2**	11,85	2	1.63	0	0	2	1.2
35	çüş*	2	11.85	**1**	5.93	0	0	1	1.6	0	0
36	iyy	3	17.78	**1**	5.93	1	0.81	1	1.6	0	0
	Total	*3954*	*19731.4*	***1305***	*7733.4*	*123*	*100*	*62*	*100*	*61*	*100*

TN: Total number of tokens in corpus, *AF:* Absolute frequency of response tokens, *RF*: Relative frequency per million,

Most frequently observed non-lexical response token is *hı-hı* 'mm-hmm' (*AF*=337, *RF*=1997.06 per million) and it is used by 64 out of 123 (52% of all speakers) speakers in the corpus. It is followed by *hıı* 'hmm' (*AF*=147, *RF*=871.12). For both of these response tokens, the number of female speakers using them is higher than the number of male speakers and both response tokens occur more in all-female conversations compared to all-male and mixed group conversations (69% of tokens of *hı-hı* and 47% of tokens of *hıı* occur in female-female talk). Excerpt (5)[27] below is from a conversation in which two male and a female 16-year-old speakers from Eskişehir talk about their high school. In turns 1 and 3, SF10016 is expressing that she is not content with the level of education they receive and the overall profile of the students in their school. While she is holding

[27] Notice that this excerpt includes the mark-up for overlaps in the speech. In the presentation of the talk, the boundaries of the overlaps are marked by < >. Please refer to **Appenfix F** to review the conventions used.

the floor, SM10001 signals that he agrees with her by inserting *hı-hı* as a response token in turn 2.

(5) Y-3-2M1F-09052021

1	SF10016	ya bak şimdi şöyle bişi var. bunu özür dileyerek söylüyorum. sakın beni yanlış anlamayın ama. şimdi hepimiz fen lisesindeyiz. yani • geldiğimizde fen lisesinde • yani açık söylemek gerekirse • çok iyi bi fen lisesi <değil! >1> ***well look, there is something like this. sorry for telling this. don't get me wrong. we are all now in a science high school. I mean, when we came to the science high school, well to be clear, it is <not>1> a very good science high school!***
2	SM10001	<hı-hı˙ >1> ***<mm-hmm. >1>***
3	SF10016	yani belki de çoğumuz LGS'de yaptığımız hatalar sonucu geldik. puanı çok düşük. baktığımız zaman. belki birileri çok çalışarak geldi. onu da bilemem. ama şimdi şey • bir fen lisesi statüsünde değiliz. bunun farkındayız hepimiz. ve bence hani bizim okuldaki insanların da • çoğunun kendini yetiştirmediğini düşünüyorum. ***I mean maybe most of us ended up here due to the incorrect question we had in the high school entrance exam. its ranking is very low. when you look at it. maybe some people got here after studying really hard. I don't know about that. but we don't have the status of a science high school. we are all aware of that. and I think most students in this school do not they are educating themselves.***

In addition to conventional non-lexical response tokens such as *hı-hı* presented previously in (5), there are also forms of interjections such as stand-alone *vay* 'wow' used as a non-lexical response token in Turkish. There are also instances of vocalizations which can be labelled as forms of slang interjections such as *oha*, *yuh*, and *çüş* (interjection *whoa* or in some cases *fuck-plus* variants can be considered as their closest English counterparts) which are observed more often as part of informal in-group conversations among (relatively younger) speakers of Turkish, compared to their use in formal registers of Turkish. Excerpt (6) below is an example to the use of interjection *oha* as a non-lexical response token.

(6) Y-2-FM-04122020

1	SF11006	ya hiç. bi de dokuzuncu sınıftayken şey var. dört • kağıtlık • sınavın • bi kağıdını direkt yapmamışım! ((laughs)) ***well, nothing. and there is this thing from the times of ninth grade. once there was an exam with four sheets filled with questions and I hadn't done one entire sheet! ((laughs))***
2	SM10004	oha! ***whoa!***
3	SF11006	ve hani ben orayı görmedim yapmadım. ***and I mean, I just didn't see that. I didn't do it.***

Though limited in terms of tokens and speakers, the data also contains tokens *wow* and *eww* as anglicisms used by Turkish speaking youth as non-lexical response tokens. In the following conversation in excerpt (7), speakers are 17-year-old classmates in an Anatolian high school in İzmir. They communicate with each other every day and the conversation takes place online. They both have the same higher socioeconomic background as the speakers in the previous conversation. The topic of their talk is cooking, SF12007 describes Spanish dessert 'churros' to her friend.

(7) Y-2-F-16122020

1	SF12007	hıı˙ belki görmüşsündür. böyle şey bi tatlı. hmm˙ biraz hamuru şey hamuruna benziyo. eklerin hamuruna benziyo ama kızartıyosun. böyle şey oluyo. tulumba tatlısı gibi gözüküyo <dışardan. >1> ***maybe you may have seen it. it is something sweet. umm its dough looks a bit like the dough of something. it looks like the dough of eclairs but you fry it. it is like this. it looks like a 'tulumba' dessert from the outside.***
2	SF12006	<hıı˙ güzeldir. >1> ***hmm it sounds good.***
3	SF12007	kızartıyorsun. böyle • ya şey şey yapıyorlar. uzun bir tane yapıp • hani böyle şu yani loop gibi yapıyolar. *(English)* böyle iki ucu böyle. kurdele gibi. ya da şey yapıyolar. çubuk çubuk yapıyolar. ***you fry it. like this. well, they make a long one. you know, they make it like a loop. it has two ends. like a ribbon. or they do it like this thing. they do it like sticks.***
4	SF12006	hı-hı˙ ***mm-hmm˙***
5	SF12007	böyle sonra • sonra onu şeker ve cinnamon'un Türkçesi *(English)* neydi? ***then, then with the sugar and what was Turkish word for cinnamon?***
6	SF12006	tarçın.

cinnamon.

7 SF12007 aynen. onun/ o ikisini karıştırıp • bir kapta. ona sıcakken ona buluyolar ya da şey yapıyorlar bulamıyolar. bir tane çikolata sosu hazırlıyolar. ona bandırıp <yiyosun. >1>
exactly. they mix those two in a cup. they coat the dough with it when it's hot or they don't coat it. they prepare some chocolate sauce. you dip it in and <eat it. >1>

8 SF12006 <wow! >1>
<wow! >1>

9 SF12007 böyle çıtır bi tatlı gibi düşün.
imagine a crispy dessert.

10 SF12006 wow that's a lot of calorie by the way.
(English) (pron. Turkish) (English)
wow that's a lot of calorie by the way.

11 SF12007 hı-hı˙ kalori evet. but who cares?
(English)
mm-hmm˙ calorie yes. but who cares?

12 SF12006 I.
English)

In excerpt (7) above, SF12006 shows her interest towards SF12006's depiction of churros with response token *wow* in turn 8. SF12007's turn is not interrupted and she continues depicting the dessert in turn 9. In this excerpt, there is also the instance of another *wow* which is not used as a response token. To elaborate, in the following turn 10, SF12006 says in English '*wow that's a lot calorie* (word *calorie* with Turkish pronunciation) *by the way*' in which *wow* now appears in a whole turn and not as a response token.

The results show that non-lexical response tokens can co-occur with other discourse particles as clusters of response tokens in youth talk. Within a L3-R3 collocation window and the minimum collocation frequency of three occurrences, a number of collocates were identified.

In Table 18 below, identified collocations for non-lexical response tokens (node token) are presented with regard to their positions to the node. The number of occurrences of each collocate is given in parenthesis. Additionally, the result showed that a number non-lexical response tokens displays reduplications, namely *ah ah*, *ay ay*, and *hı-hı hı-hı*.

Table 18 Collocates and reduplications of non-lexical response tokens

Collocations			*Reduplications (type)*
Collocate left-hand	*Node token*	*Collocate right-hand*	
-	**ah**	be (6)	ah ah
aga (11) ah (6) yok (5) vay (5)	**be**	-	-
-	**ha**	evet (9) tamam (7) anladım (6) iyi (4) doğru (3)	-
-	**hı**	evet (5) aynen (3)	-
-	**hıı**	anladım (4) işte (3)	-
-	**hı-hı**	evet (4) biliyorum (3)	hı-hı hı-hı
-	**hmm**	anladım (4)	-
-	**vay**	be (5) anasını (3)	-

As the table above shows, non-lexical response tokens co-occur with other tokens which can either be other non-lexical response tokens such as *ah be* or in the majority of cases lexical response tokens such as *ha evet* 'oh yeah'*, hı aynen* 'ah exactly'*, hı-hı biliyorum* 'mm-hmm I know'. In the following section, these lexical response tokens will be presented and exemplified in detail.

4.3.1.3.2 Lexical response tokens

A total of 1728 lexical response tokens of 37 types were identified in the corpus. In Table 19 below, total number of lexical response tokens tabulated by speakers are presented. Total number of tokens (*TN*) retrieved from the corpus is presented and the total number of identified response tokens (*AF*) along with the relative frequencies (*RF)* per million are presented in a descending order. For each lexical response token, total number of unique speakers, number of female and male speakers are also presented to illustrate the extent each token is used by Turkish speaking youth in the study.

Table 19 Lexical response tokens tabulated by frequencies and speakers

Rank	*Lexical response token (type)*	*Gloss*	*No. of tokens*				*No. of addressers*					
			All occurences		*Lexical res. tokens*		*All*		*Female*		*Male*	
			TN	*RF*	*AF*	*RF*	*N*	*%*	*N*	*%*	*N*	*%*
1	evet	*yes*	1582	9374.93	**793**	4699.31	97	79	55	89	42	69
2	aynen	*exactly*	656	3887.45	**329**	1949.65	80	65	40	65	40	66
3	hayır	*no*	301	1783.72	**78**	462.23	40	33	20	32	20	33
4	yok	*negative existential exp.*	451	2672.62	**65**	385.19	40	33	15	24	25	41
5	anladım	*I see*	106	628.16	**50**	296.30	28	23	17	27	11	18
6	tamam	*okay*	776	4598.57	**44**	260.74	29	24	16	26	13	21
7	bence de	*I think so too*	76	450.38	**36**	213.34	25	20	15	24	10	16
8	iyi	*good*	621	3680.04	**35**	207.41	26	21	12	19	14	23
9	öyle	*so*	782	4634.13	**32**	189.63	21	17	10	16	11	18
10	yani	*well*	2126	12598.67	**26**	154.08	17	14	12	19	5	8
11	olabilir	*it might be*	246	1457.80	**25**	148.15	19	15	12	19	7	11
12	doğru	*correct*	160	948.16	**22**	130.37	16	13	6	10	10	16
13	işte	*well*	881	5220.80	**17**	100.74	16	13	12	19	4	7
14	ciddi misin	*are you serious?*	17	100.74	**16**	94.82	8	7	5	8	3	5
15	tabii	*of course*	118	699.27	**14**	82.96	22	18	3	5	19	31
16	valla	*really (lit. I swear)*	180	1066.68	**14**	82.96	14	11	3	5	11	18
17	Allah(ım yarabbim)	*God/Oh my God*	211	1250.39	**13**	77.04	8	7	5	8	3	5
18	harbi (mi)	*really (?)*	48	284,45	**12**	71.11	6	5	2	3	4	7
19	neyse	*anyways*	165	977.79	**10**	59.26	9	7	5	8	4	7
20	di mi	*truncated form of tag question*	307	1819.28	**9**	53.33	9	7	5	8	4	7
21	ne	*what*	928	5499.32	**9**	53.33	2	2	2	3	0	0

Table 19 (cont'd)

Rank	*Lexical response token (type)*	*Gloss*	*No. of tokens*				*No. of addressers*					
			All occurences		*Lexical res. tokens*		*All*		*Female*		*Male*	
			TN	*RF*	*AF*	*RF*	*N*	*%*	*N*	*%*	*N*	*%*
22	olur	*that's fine*	104	616.30	**9**	53.33	9	7	7	11	2	3
23	peki	*alright*	48	284.45	**9**	53.33	8	7	6	10	2	3
24	inşallah	*hopefully (lit. God willing)*	48	284.45	**8**	47.41	8	7	4	6	4	7
25	maşallah	*wonderful (lit. by God's will)*	17	100.74	**8**	47.41	5	4	3	5	2	3
26	anasını +	*mother-plus swear exp.*	54	320	**7**	41.48	5	4	1	2	4	7
27	yazık	*pity*	29	171.85	**7**	41.48	7	6	4	6	3	5
28	eyvallah	*thanks*	12	71.11	**6**	35.56	5	4	0	0	5	8
29	okey	*okay*	59	349.63	**5**	29.63	4	3	2	3	2	3
30	siktir +	*fuck-plus swear exp.*	28	165.93	**5**	29.63	5	4	1	2	4	7
31	aman	*oh my (lit. caution)*	20	118.52	**3**	17.78	3	2	0	0	3	5
32	biliyorum	*I know*	131	776.31	**3**	17.78	3	2	2	3	1	2
33	amına +	*vagina-plus wear exp.*	152	900.75	**2**	11.85	2	2	0	0	2	3
34	kesinlikle	*absolutely*	41	242.97	**2**	11.85	2	2	2	3	0	0
35	muhtemelen	*probably*	31	183.71	**2**	11.85	2	2	2	3	0	0
36	tövbe	*exp. for disbelief/disapproval*	9	53.33	**2**	11.85	2	2	2	3	0	0
37	şaka	*joke*	28	165.93	**1**	5.93	1	1	1	2	0	0
	Total		*811*	*4805.983*	***79***	*468.15*	*123*	*100*	*62*	*100*	*61*	*100*

TN: Total number of tokens in corpus, *AF:* Absolute frequency of vocatives, *RF*: Relative frequency per million

Token *evet* 'yes' is among the most frequently occurring words in both general spoken Turkish and Turkish youth language. It ranks 17th both in the CoTY and the STC wordlists (*AF*=1582, *RF*=9374.93 per million; *AF*=2249, *RF*=7852.90 per million, respectively). As a lexical response token as well, it is the most frequently occurring (*n*=793, 45% of all lexical response tokens) token used by the vast majority of speakers (*n*=97, 79% of all speakers) in the corpus. Following *evet*, the most frequently occurring lexical response token in Turkish youth language is *aynen* (*AF*=329, *RF*= 1949.65) which is previously reported as a backchannel with a very infrequent use (*AF*=3, *RF*=19.93) in general spoken Turkish (Aytaç-Demirçivi, 2021). In Turkish youth language though, it is the second most frequently used lexical response token. It is also interesting to note that number of occurrences for each of the remaining response tokens are drastically lower, *evet* and *aynen* make up of 64% of all lexical response tokens in the corpus. In excerpt (8) below, both *evet* 'yes' functions as response token while *aynen* 'exactly' does not as it holds a turn.

(8) Y-2-F-14052021-2b

1	SF11011	o yüzden • böyle bi giyilecek şeyler var. bi de daha fotoğraflık • olan şeyler var. ***that's why there are things to wear like this. and there are things for taking photos.***
2	SF09007	*aynen*. ben mesela fotoğraflık olan şeyleri böyle yani çok • almam. yani giymem. ***exactly. for example, I do not buy things which are for photos. I mean I don't wear them.***
3	SF11011	ben giyiyorum. ***I do.***
4	SF09007	çünkü her zaman giyebileceğim şeyler olmuyo. ***because those things are not the kind of things that I can wear all the time.***
5	SF11011	bazen gerçekten •influencer olmanın faydalı olacağını düşünüyorum. mesela • e˙ dün konuştuk ya. ‿bizim hiç böyle fotoğrafımız yok diye. ‿fotoğraflara <bakarken.>1> ***sometimes I am thinking that being an influencer would be beneficial. for example, remember what we talked about yesterday about how we don't have any photos like those while we are <looking at>1> the photos.***
6	SF09007	<evet.>1> ***<yes. >1>***
7	SF11011	mesela onların • ilerde çok fazla anısı olmuş olcak. çünkü • her adımlarını kaydediyolar. ***well, they will have so many memories in the future. because they are recording their every step.***
8	SF09007	evet. ***yes.***

In the excerpt above, SF11011 is talking about influencers who share their personal life on social media and SF09007 uses *aynen* in her first turn (turn 2) and then two instances of *evet* in her following turns (turns 6 & 8) to signal to her friend that she is listening to SF11011 and wants her to keep talking. Both response tokens semantically have affirmative associations such as agreement and continuation yet their functional scope needs further analysis.

The identified lexical response tokens also include expressions which are in the form of questions. These include follow-up questions *ciddi misin* 'are you serious?' (*n*=16), *harbi mi* 'really?' (*n*=12), and tag question *di mi* 'isn't it?' (*n*=9) which do not expect responses but rather indicate engagement and attention of the listener. In (9) below, the response token *di mi* which is the truncated form of tag question *değil mi* in informal spoken Turkish is used by a 18-year-old female speaker in an episode of gossip talk[28]. This tag question is a combination of negative particle *değil* (lit. not) and *mI* (clitic used to form questions in Turkish) which in general corresponds to *isn't it* in English. The topic of the talk is a mutual friend who is cohabiting with her boyfriend. SF11005 judge this situation as a morally inappropriate behaviour and in turn 4, SF12005 uses *di mi* as an encouragement for SF11005 to continue to talk.

(9) Y-2-F-24122020

1	SF11005	kanka bence/ bence babası bilse bile çocuklarını falan bilmiyodur. kesinlikle. ***kanka, I think even if her father knows, the children do not know it. definitely.***
2	SF12005	haˑ onu bilmiyorum. ***well I don't know about that.***
3	SF11005	annesi acaba biliyo mu acaba çocuklarını? ((1.0)) kanka bi aile nası/ bak gerçekten. yani bak ben sadece şey olarak düşünüyorum. bi aile buna nası izin veriyo? ***I wonder whether her mom knows about his children? ((1.0)) kanka what kind of a family would –I mean, look, I am just thinking, how does a family allow that?***
4	SF12005	di mi? ((fast)) ***isn't it?/right?***
5	SF11005	verir misin sen? hayır! ***would you? no!***

It is observed that some of the identified lexical response tokens can appear in clusters as well. In Table 20 below, within a L3-R3 collocation window and the minimum

[28] Following Coupland (2003), Coupland & Jaworski (2003) and Jaworski & Coupland's (2005) works, this study defines **gossip talk** as episodes of small talk among friends who engage in (usually pejorative) evaluative conversation about people and their personal lives.

collocation frequency of three, collocations for lexical response tokens (node token) which appear in the left context and right context as well as reduplication forms are presented. The number of occurrences of each collocate is given in parenthesis.

Table 20 Collocates and reduplications of lexical response tokens

Collocations			*Reduplications (type)*
Collocate left-hand	*Node token*	*Collocate right-hand*	
ha (6)	**anladım**	-	anladım anladım
hı (3) hı-hı (3)	**aynen**	öyle (13)	aynen aynen
evet (4)	**bence de**	-	-
ha (5)	**doğru**	-	doğru doğru
ha (12) yani (5) hı-hı (5) yani (5) hı (5)	**evet**	ya (11) aynen (6) öyle (5) bence de (4) abi (3) gerçekten (3)	evet evet, evet evet evet
-	**hadi**	ya (5)	hadi hadi
-	**harbi**	mi (6)	-
-	**hayır**	ya (4)	hayır hayır
tamam (3)	**işte**	-	-
çok (10) ha (4)	**iyi**	-	iyi iyi
aynen (11) valla (3)	**öyle**	-	-
-	**tabii**	canım/oğlum/lan (4)	tabii tabii
e (3) ya (3) ha (4)	**tamam**	-	tamam tamam
cık (7)	**yok**	kanka/oğlum/la (13) ya (10) be (5)	yok yok

The table above illustrates that both lexical response tokens such as *evet* 'yes' and non-lexical response tokens such as *cık* (see section 4.3.1.3.1 for the overview of non-lexical response token in the corpus) co-occur with the lexical response tokens listed. Address terms such as *kanka* 'dude', *oğlum* (lit. my boy, closest equivalent would be 'dude'), *lan (*a derivative of *oğlan* 'boy'*)* exhibit collocational relationship with lexical response

tokens (see 4.3.2 for more details on vocatives in the corpus) as well. Response token *anladım* 'I understand', *aynen* 'exactly', *doğru* 'right', *evet* 'yes', *hadi* 'come on', *hayır* 'no', *iyi* 'good', *tabii* 'of course', tamam 'okay', and *yok* (negative existential expression) are used in reduplicated forms which suggest intensification of the pragmatic function achieved by these markers. In excerpt (10) below, an example for lexical response token reduplication is presented.

(10) Y-2-F-21122019

1	SF09005	ya aslında ben bu ara şeye gitcem. ya kanka • yüzmeye gitcem. gelir misin? takılalım. ***by the way, I'll go to that thing. kanka, I'll go to the swimming pool. would you come? we can hang out.***
2	SF09006	takılalım bro valla. ***let's hang out, bro.***
3	SF09005	sen şimdi dersin • yine şey sınav haftası falan. <sıkıntı falan edersin.>1> ***but you could say –again- that it is the midterms week or something. <that would be a problem for you or something .>1>***
4	SF09006	<yok yok. >1> ***<no no. >1>***
5	SF09005	sıkıntı yoksa sıkıntı vardır. ***there is a problem if there is no problem.***

The excerpt above is from a face-to-face conversation from Elazığ, a 15 year-old female is inviting her friend to a swimming pool. In turn 3, SF09005's utterance *sen şimdi dersin yine şey sınav haftası falan* 'but you could again say it is the midterms week and stuff' is a reference to their earlier conversation at the very beginning of the recording where SF09006 indicated that she preferred to spend her spare time studying for school rather than engaging in leisure or social activities. Thus in turn 4, SF09006 uses the reduplicated response token *yok yok* 'no no' to indicate that she genuinely accepted this offer in her previous turn in 2.

4.3.1.3.3 Unconventional forms: Words of slang, taboo, religion

Apart from the conventional forms of response tokens identified in the existing literature for spoken Turkish, the results also reveal the tokens within the domains of taboo language, religious discourse and contemporary slang which are also used as response tokens by Turkish speaking youth in the corpus. In this line, the results corroborate what was reported previously for British English, Irish English (Adolphs & Carter, 2013; Drummond, 2020; Love, 2017; Stenström, 2017) and Spanish (Amador Moreno et al.,

2013) which also utilized religious references and swear words also act as response tokens.

In relation to register of religious discourse, the corpus includes words and expressions *valla*[29] (*n*=14)*, Allah* (*n*=13)*, inşallah* (*n*=8)*, maşallah* (*n*=8), *tövbe* (*n*=2) which are all originally Arabic. Among them *Allah* appears in reduplicated forms. Excerpt (11) is an example to this where SF11013 is explaining her friend that the pool in her family house requires too much work and SF11012 is using different forms of lexical response tokens such as *evet* 'yes' in turn 2 (co-occurring with *işte*), and the subsequent *Allah Allah* 'Oh my God' in turn 4, neither of which takes the speaker turn.

(11) Y-2-F-05062021

1	SF11013	abi havuz aşırı zahmetli bi işmiş! ***man, the pool is such a hassle!***
2	SF11012	evet işte. ***oh yeah.***
3	SF11013	aşırı yani! yok onun gideri var. temizlenmesi var. tekrar temizlenip havuza pompalanması var. yaklaşık kaç ton mu? on ton mu? yüz ton mu? on ton değildir. yüz ton su alıyo herhalde şu an orası. ***so much! there is the drainage. the cleaning. flushing and pumping to the pool. how many tons approximately? ten tons? a hundred tons? not ten tons. I guess it takes about a hundred tons of water right now.***
4	SF11012	Allah Allah! ***Oh God!***
5	SF11013	çok fazla. hayır ee˙ daha güzel bi şey var. acaba biz yüz ton suyu nerden bulcaz? ***too much. no well there is something even better. I wonder where we will find a hundred tons of water?***

The group of swear words used as response tokens include clusters of expressions derived from node words *anasını* (*n*=7) which corresponds to mother-plus swear words, *siktir* (*n*=5) which is equivalent to *fuck* and its variants, and *amına* (*n*=2) which encompasses vagina-related swearwords in Turkish. These words and expressions are marked as having taboo value in Turkish language. Both male and female speakers make use of them as response tokens as in excerpt (12) below where response token *vay anasını* (truncated form of mother-plus swearing expression preceded by another response token *vay*) is used by SF10012 in turn 2.

[29] truncated form of *vallahi*

(12) Y-2-F-18052021

1	SF10011	hayır lan! o şey • babamın Facebook'unda paylaştık. paylaşmadığı için otuz takipçisi falan vardı. Facebook'ta duyurduk. herkes bizim <reklamımızı yapıyo. >/1>. ***no man! that thing, we shared it on my dad's Facebook. he had thirty followers or something because he was not sharing anything. we announced it on Facebook. everybody is promoting us.***
2	SF10012	<vay anasını! >/1> <***wow holy mother***! **>/1>**
3	SF10011	teyzem • işte Antep'e gelin giden. gruplarına atmış. zaten kuzenleri falan üye olduğu için otuz tanesi falan birden geldi. ***my aunt, the one who married into a family from Antep. she shared it in their (Facebook) groups. thirty of her cousins subscribed to us at once.***
4	SF10012	((laughs))

As presented in (12) above, SF10012's swearing response token also overlaps with a part of SF10011's utterance in turn which displays the high level of engagement in interaction achieved through this particular response token. As one of the speakers is telling a story, the listener conveys the message that she is interested in the course of events narrated by the speaker. The use of swearing expression *vay anasını!* 'wow holy mother!', in this context then, marks a shared emotive stance. It is suggested that this alignment facilitates the ingroup bonding as was also observed for the swearing practices among British youth previously (Drummond, 2020).

Lexical response tokens also includes lexis from contemporary slang such as *harbi* 'really' (*n*=12)*, aga be* 'come on bro' (*n*=10)*, şaka* 'joke' (*n*=1) which do not appear as response tokens in general spoken language of Turkish represented by the STC data[30]. Response token *aga be* is a contemporary example for slang expressions used as response tokens in the CoTY. This expression consists of the non-lexical response token *be* and its collocate aga (derivative for elder brother *ağabey* in Turkish) and occurs as a response token 10 times by 7 unique speakers (4 females and 3 males) in the corpus. The results show that this particular response token is used to express emotive engagement by Turkish speaking youth in the CoTY.

[30] based on separate analysis conducted on the STC-Beta version of 350,000-words (Ruhi et al., 2010). Access granted by Dr. Hale Işık-Güler.

In excerpts (13) and (14) below, speakers use *aga be* to convey emotions, anger in (13) and empathy in (14), to respond to the stories their interlocutors are narrating.

(13) Y-3-2M1F-14052021

1	SF11008	o ((name_female)) yüzünden • sırada oje kazıdığımı hatırlarım. ***because of that ((name_female)), I had to scratch my nail polish.***
2	SM11004	aga be! ***aga be!***
3	SF11008	iki dakika içinde. ***in two minutes.***

To elaborate, in excerpt (13), which is an online conversation between 17-year-old female and male speakers from Çanakkale, the speakers are criticizing the strict school management regulations deployed by the headmistress of their high school. SF11008 recalls a time when she had seen the headmistress approaching and had to instantly remove her nail polish in order not to get punished. In turn 2, SM11004 responds to his friend with the response token *aga be* in order to demonstrate his anger towards the headmistress for causing his friend stress back then.

In (14) below, response token *aga be* is used in s a face-to-face conversation between two 17-year-old female speakers from Ankara. In the conversation, SF10012 implies that she likes a raincoat her friend SF10011 has. When SF10012 inquires about the raincoat, SF10011 shares that it is actually not an original product and that was why she did not want to wear it. In turn 3, as a reponse, SF10012 replies with *aga be* to express her emotive engagement with her friend.

(14) Y-2-F-18052021

1	SF10012	şey bu yağmurluk olarak mı geçiyo? ***is this called a raincoat?***
2	SF10011	bilmiyorum. üstünde Adidas yazıyodu. ama burda da Nike yazıyo. o yüzden giyiyim demedim. ***I don't know. it said Adidas on it. but here it says Nike. that's why I didn't want to wear it.***
3	SF10012	aga be! ***aga be!***
4	SF10011	şimdi her tarafına marka yapıştırması ile ilgili özentisi. ((laughs)) ***now it's about that wannabe branding all over the place. ((laughs))***
5	SF10012	((laughs))

The lexical item *okey* is an anglicism of *okay* which occurs 59 times in the CoTY and among them, 5 of the instances act as response tokens used by both males and females. Below excerpt (15) is an example of this particular response token in a conversation among three 18-year-old male friends from Mersin. The topic of the talk is action movies and SM12012 is recommending an action movie to his friends. In the talk, SM12013 fills the response slots with lexical and non-lexical response tokens without claiming the turn. In his first turn (turn 2), a combination of non-lexical and lexical response tokens *haa öyle* (i.e. *haa* 'oh' and *öyle* 'I see') is present as a pair, and the following turn (turn 4) *okey* is used in an extended response. Though it is an extended turn, the turn is still not yielded.

(15) Y-3-M-06122020-2

1	SM12012	süper kahraman değil ya. böyle doğaüstü bişey yok. bildiğin bi asker • adam. ***not a superhero. there's nothing supernatural like that. he's just a soldier. a man.***
2	SM12013	haa˙ öyle. ***oh I see.***
3	SM12012	ama doğaüstü yok derken • biraz John Wick gibi. ***though not supernatural, a bit like John Wick.***
4	SM12013	ha˙ bayağı. okey. ***ah like that. okay.***
5	SM12012	vurdulu kırdılı. adam • one man army gibi bişey yani. (English) ***it's blood-and-guts. It's like one man army or something.***

The use of both conventional forms such as *evet* 'yes' as well as non-conventional forms such as *şaka* 'joke' indicate that speakers feed their linguistic repertoire from various sources and such richness lays ground for their multi-functionality in interaction. In order to explore the various pragmatic functions in co-text and context, the following section will focus on a particular lexical response token *aynen* 'exactly' used in Turkish youth talk.

4.3.1.3.4 Register-specificity: *aynen*

Aynen is defined as an adverb by the official Turkish Language Association (TDK) Dictionary and frequency dictionary of Turkish (Aksan et al., 2017) in the sense of 'just like that, as it is'. Stand-alone *aynen,* however, appears to have an additional meaning not

presented by the above-mentioned sources. As contrasted in examples from *written* Turkish data from the TNC in (i) and *spoken* Turkish data from the STC in (ii) below, stand-alone *aynen* does not function as an adverbial modifying other parts of speech in spoken Turkish, rather in (ii), its closest equivalent in English would be *exactly*.

(i) Ben de ona aynen bana vurduğu gibi vurdum.
'*I hit him just like how he hit me.*'
[source: TNC-V.03, W-RA16B1A-1213-2219]

(ii) CEV000041: ee aileye alıyordunuz yani.
'*so you were buying for the family.*'
HAS000795: aynen. ((0.3)) dokuz yüz kilo babam şey aldı.
'*exactly. ((0.3)) my dad bought nine hundred kilos of that stuff.*'
[source: STC-Beta, 072_111017_00300]

In the CoTY, the KWIC analysis shows that lexical token *aynen* usually appears in stand-alone position in utterances (*n*=474, 72% of total occurrences, reduplications included) which suggests that its primary function in dyadic or multi-party talk is not an adverbial as prescribed in its dictionary meaning. *Aynen* remains scarcely explored within Turkish linguistics, the only study which mentions this token is Kaynarpınar's (2021) corpus-driven analysis in which *aynen* is defined as an agreement marker used to indicate total approval on an opinion or a suggestion. Below are sample concordance lines for *aynen* in stand-alone position in the CoTY.

Left Context ▲	Match	Right Context
	aynen	ben de onu izliyorum. eleştiri kanalı. ben de onu
	aynen	Minecraft videoları falan çekiyo. ben de onu izli
	aynen	.
	aynen	aynen. o notayla falan da uymuyo. çok rahatsız ed
	aynen	aynen. Carry Potter. o güzel. ((1.0)) ıı ˙
	aynen	Porçay'ı ben de izliyom. Porçay. Kaanflix. onlar
	aynen	.
	aynen	. evet tam olarak big boyfriend bulmaya gelmiş. ba
	aynen	aynen. ona benziyo hafiften. o yüzden hoşuma gidi
	aynen	. onun/ o ikisini karıştırıp bir kapta. ona sıcakk
	aynen	aynen aynen aynen! hani deli gibi kalın/ pardon k
	aynen	((laughs)). ee ˙orda falan şey • yani çok mesela z
	aynen	((2.0)) bilmiyorum ya! öyle işte. ne diyebilirim
	aynen	. ama işte sonradan • işler biraz da sarpa sardı.
	aynen	!

Figure 13 Sample concordance lines for *aynen* in the CoTY

It is important to note that while lexical token *aynen* ranks 42nd in the wordlist of the CoTY (*AF*=656, *RF*=3887.45), it is relatively quite low in frequency in the TNC (*AF*=3387, *RF*=66.83) and in the STC (*AF*=56, *RF*=195.53). The STC covers the period of 2008-2013 and mainly has speakers above 18 years old while the data of the CoTY was collected ten years later and the ages of speakers are between 14 to 18. In this line, the discrepancy between the frequencies of occurrences of *aynen* between two corpora might either suggest a recent trend in spoken Turkish and/or identify *aynen* as a register-specific lexis for Turkish youth language.

The results show that *aynen* stands out as an extensively used response token in Turkish youth interaction in the CoTY. It is the second most frequently used lexical response token which corresponds to 19% of all lexical response tokens in data (*AF*=329, *RF*=1949.65 per million) and it is used by the majority of speakers (*n*=80, 71%). The distribution of *aynen* according to age of speakers (see Table 21 below) shows that speakers from all ages use this token, with 17-year-olds being the most frequent users of *aynen* as a response token in their speech.

Table 21 Distribution of *aynen* by speaker age

Age	*No. of speakers who used aynen*	*% of speakers in age group*
14	9	75
15	5	45
16	30	74
17	19	87
18	17	71
Total	80	

Response token *aynen* is extensively used by speakers across all socioeconomic backgrounds (a balanced distribution is observed across main socioeconomic groups; LOW 35%, MIDDLE 32.5%, HIGH 32.5%), enrolled in all types of schools, and from 93% of the reported cities of residence in the CoTY.

With regard to conversations in which *aynen* is used as a response token (*n*=43, corresponds to 83% of the total number of conversations in the corpus), groups which only had female speakers used *aynen* more frequently (*n*=172, %52 of all occurrences)

than all male and mixed groups in the CoTY. The distribution of occurrences based on speaker groups is elaborated in Table 22.

Table 22 Distribution of *aynen* tabulated according to speaker groups

Speaker group	*No. of conversations*		*No. of speakers*		*No. of occurences*	
	N	*%*	*N*	*%*	*N*	*%*
all female	19	44	33	41	172	52
all male	14	33	28	35	79	24
mixed	10	23	19	24	78	24
Total	*43*	*100*	*80*	*100*	*329*	*100*

For the purposes of identifying functions of *aynen* in Turkish youth talk, the taxonomy for categorizing listener response tokens developed by O'Keefe and Adolphs (2008) was utilized which takes a discourse and pragmatic perspective on the response tokens in casual conversation and consists of both lexical and non-lexical responses. This analytical framework was adopted due to the fact that the register, methodological tools and analytical perspectives are compatible with the present study.

O'Keeffe and Adolph's (2008) taxonomy offers four broad categories of functions which are (i) *continuers* (after Schegloff, 1982) (i.e. marking expectation for the speaker to continue), (ii) *convergence tokens* (i.e. marking points of topic change/shift), (iii) *engagement tokens* (i.e. marking emotive response to the message), and (iv) *information receipt tokens* (i.e. marking that adequate info has been received). Though the taxonomy emerged from British and Irish English data, a successive study has adapted it for Spanish response tokens as well (Amador Moreno et al., 2013).

Within this taxonomy, the results show that *aynen* is most frequently used as a *continuer* (47%) followed by the function of *convergence* (25%) and then *engagement* (15%). The smallest proportion of functions (13%) belongs to the group of *information receipt* (see Table 23).

Table 23 Functional distribution of *aynen*

Function	*No. of occurences*	*% of functions*
continuer	141	47
convergence	75	25
engagement	46	15
information receipt	35	13
Total	*297**	*100*

*reduplications counted as a single cluster

The most frequently used function of *aynen* is ***continuer*** (42%) which was originally identified by Schegloff (1982) as the function of active listenership. As a continuer, speakers use *aynen* to maintain the flow of discourse and encourage the current speaker to continue to talk as exemplified in excerpt (16) below. The excerpt is from an online conversation between a 16-year-old male and 17-year-old female from İzmir. Male speaker SM10004 pours out his heart to his close female friend SF11006 that he is concerned with the closeness his girlfriend displays towards another male classmate. While he is narrating an episode of such affectionate behavior, SF11006 uses *aynen* to signal her interest in the topic and show desire for SM10004 to continue telling the story.

(16) Y-2-FM-04122020

1	SM10004	o kadar da değil! ama şey yani • ee˙ mesela diyelim biz yürüyoruz • tamam mı? ***it' not that much! but, I mean, well, let's say we are walking, okay?***
2	SF11006	hı-hı˙ ***mm-hmm˙***
3	SM10004	bak. biz varken. diyelim işte sarıldık • ettik. yürümeye başladık. bir anda mesela ((name_male))'ın koluna girip yürüyebiliyo anladın mı? ***look. when we were together. let's say we hugged and stuff. we started walking. suddenly, for example, she just takes ((name_male))'s arm and walks, you know?***
4	SF11006	hıı˙ ***yeah.***
5	SM10004	ve ben orda normal yürüyorum falan. bu • mesela sinirimi bozuyo. niye benle yürümüyosun? çünkü • çok sık yaptığımız bişey de değil. ***and I am just walking casually there and stuff. this, for example, gets on my nerves. why don't you walk with me? because this is not even something we do very often.***
6	SF11006	aynen. ***exactly.***

7	SM10004	yani bu tarz şeylere tilt oluyorum. bir de ((name_male))'a yapması ekstra oluyo. çünkü ((name_male))'ın da geri basmadığını biliyorum bu konuda. ***I mean I can't stand these kind of things. and doing this to ((name_male)) is just too much. because I know ((name_male)) does not hold back, either.***

In the excerpt, SF11006 displays her active listenership through non-lexical response tokens *hı-hı* in turn 2 and *hıı* in turn 4 both of which also function as continuers. SM10004 continues complaining and in the following turn of 5, SF11006 uses token *aynen* to encourage her friend to carry on in turn 6. It is important to highlight that aynen is used not as a response to SM10004's statement *'(it) gets on my nerves'* but as a response to his ongoing re-enactment '*why don't you walk with me? because this is not even something we do together often'* as a whole in turn 5. The excerpt shows that the use of *aynen* does not interrupt SM10004's turn but rather facilitates the expansion of the narrative, SM10004 continues with his talk without any disruption in turn 7.

In Turkish youth interaction, continuer *aynen* also functions as a marker in the co-construction of interactional humour. Excerpt (17) is an unfolding episode of conversational humour manifested among two 18-year-old male classmates from a science high school in Mersin.

(17) Y-2-M-03122020

1	SM12008	((name_town))'ta havalar nasıl? ((laughs)) ***how is weather like in ((name_town))? ((laughs))***
2	SM12009	((laughs)) havalar güzel. ((3.0)) iyi. ***((laughs)) weather is good. ((3.0)) nice.***
3	SM12008	yok ya. ben soğuğu sevmiyom. ***nah. I don't like the cold.***
4	SM12009	sen yaz adamı mısın? ***are you a summer guy?***
5	SM12008	yo! ***nope!***
6	SM12009	yaz aşkı. yaz aşk. ***summer love. summer. love.***
7	SM12008	ben kendimi mevsimlere göre değerlendirmem de. ((laughs)) ***I don't define myself by seasons anyway. ((laughs))***
8	SM12009	kendini neye göre değerlendirirsin? ((laughs)) ***what do you define yourself by? ((laughs))***
9	SM12008	kendimi • bilmiyom ya. ben • sistem adamı oldum. ben bu seneyi sistem adamı olarak geçiricem.

		I don't know. I've become a man of the system. I'll spend this year as a man of the system.
10	SM12009	ya en kârlısı o <zaten. >1> ***well that's the most profitable thing <anyway. >1>***
11	SM12008	<aynen. >1> ∪aynen. ***<exactly. >1> ∪exactly.***
12	SM12009	bunu er ya da geç fark edeceğiz. ***we will realize this sooner or later.***
13	SM12008	ama bi senelik. seneye ben yoluma devam etmeyi düşünüyom. ((short laugh)) tabii mezun tayfaya katılmazsam. ***but it is only for a year. next year I am thinking of moving on. ((laughs)) unless I join the crew of grads.***

In the excerpt, the episode of humour starts unfolding in turn 7 through SM12008's statement *ben kendimi mevsimlere göre değerlendirmem de* 'I don't define myself by seasons anyway' *which* is followed up by SM12009's question *kendini neye göre değerlendirirsin?* 'what do you define yourself by?' to expand the narrative. Notice that both speakers laugh after these utterances, which signals a playful tone in the exchange. In the following turn, SM12008's reply is a form of self-deprecating humour (Dynel, 2008, 2009; Norrick, 1993, 2009; Lampert & Ervin-Tripp 2006) through the utterance in turn 9 *ben sistem adamı oldum* 'I've become a man of system' which displays a pseudo self-putdown with the purpose of triggerring a common theme in the sociopolitical space speakers belong to, thus building solidarity. This self-deprecating utterance elicits SM12009's engagement in the common theme and he collaborates to expand the humourous narrative with the expression *ya en kârlısı o zaten* 'well that's the most profitable thing anyway' in turn 10. At that point, in turn 11, SM12008 uses duplicated form *aynen aynen* 'exactly exactly' to signal his interlocutor to continue, and SM12009 holds the floor for another turn. Later SM12008 takes the turn and concludes the episode of humour accompanied by laughter which establishes a tone of sarcasm in the utterance as was previously observed for spoken Turkish (Uçar & Yıldız, 2015).

The second most frequently observed function for *aynen* is ***convergence***. Adolphs and Carter (2013, p. 53-54) note that convergence tokens operate on points of talk when speakers "agree, or simply converge on opinions or mundane topics and this leads them collaboratively to negotiate topic boundary points, where a topic can be shifted or changed". Similarly, in interaction among Turkish speaking youth, convergence tokens are used for shifting the topic as in (18) where three 16-year-old female friends start

talking about politics but then SF10009 does not feel comfortable talking politics on record. SF10008 acknowledges SF10009's concern and proposes to talk about the issue some time later in turn 4. SF10009 uses *aynen* to converge with SF10008 and mark the topic closure. Following that in turn 6, SF10008 launches a totally different topic.

(18) Y-3-F-14052021

1	SF10008	ne güzel sallıyolar ha! ***they are talking crap!***
2	SF10009	((laughs)) siyaset yapmayalım! ***((laughs)) let's not do politics!***
3	SF10010	ben de… ***me too…***
4	SF10008	((laughs)) tamam sustum. bunu seninle ayrıca başka zaman <konuşuruz.>1> ***((laughs)) okay I am not saying anything else. we can talk about this with you <later.>1>***
5	SF10009	<aynen.>1> ***<exactly. >1>***
6	SF10008	şimdi hazır mısın? yine çöpe iniyoruz. ama araba geliyo. ama amca bi dur ya! ***now, are you ready? we are going out to throw garbage again. but a car is coming. but but uncle, stop!***

The analysis also indicate that when speakers use *aynen* to show agreement, the convergence token is also oriented around *face* concerns[31] (Brown & Levinson, 1987; Goffman, 1967). In line with Adolphs and Carter's (2013) note on the high relational value of convergence tokens, young speakers of Turkish make use of response token *aynen* to converge in order to avoid face threats and maintain good relationship between their interlocutors. As an example, excerpt (19) below presents an instance of this function. Speakers are three 18-year-old male speakers who are classmates for three years. They all live in Mersin, attend an Anatolian high school and are from lower socioeconomic backgrounds. The main topic of the talk is the shows they watch. SM12014 is an anime fan while the other two do not watch anime or any animated series. Prior to the excerpt presented here, SM12012 and SM12013 argue that anime as a genre

[31] Brown and Levinson's (1987) built their Politeness Theory on the notion of **face** initially defined as "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact" previously by Goffman (1967, p. 5). Politeness Theory assumes that there are two components of face: **positive face** which refers to a person's desire to have a positive self image and **negative face** which refers to a person's need to be independent. In interaction, the interactants may encounter **face threatening acts** to their positive and negative face.

cannot be as good as other forms of visual media. Upon one of the speaker's mention of 'Avatar: the Last Airbender' which is an American animated series, SM12012 states that he does not have any interest in that show, either. SM12014 shows a sign of disagreement with SM12012's downgrading of the series and shares his positive opinion regarding the series in turn 3. SM12012 justifies that the series 'didn't hook him in' in turn 4. In the following turn, due to the fact that SM12014's repetitive attempt to change his friend's negative opinion is an imposition, SM12012 responds with a brief and direct utterance to display opposition *kötü demedim* 'I didn't say bad' followed by *hayır* 'no' in turn 6. So as not to further threaten his interlocutor's face by imposing and also avoid face threatening acts on his positive face as well, SM12014 uses *aynen* as a mitigator.

(19) Y-3-M-06122020-2

1	SM12012	Avatar izlerken de mesela öyle hissetmiştim. ***that's how I felt when I watched Avatar.***
2	SM12013	ama Avatar güzeldi. ***but Avatar was good.***
3	SM12014	niye • Avatar güzeldi yani. ***why, Avatar was good.***
4	SM12012	beni bağlayamamıştı. yani çekememişti. ‿yani bıraktım o yüzden. ***it didn't hook me in. it didn't draw my attention. I mean that's why I quit.***
5	SM12014	yine yine çok kötü değildi yani. iyi yani o. ***still still it wasn't so bad. it is good I mean.***
6	SM12012	kötü demedim • hayır. ***I didn't say bad, no.***
7	SM12014	aynen. ***exactly.***
8	SM12012	bana hitap etmedi ama. bağlayamadı beni kendine. ***but it didn't appeal to me. it couldn't pull me in.***
9	SM12014	*((name_male))*• Allah'ı • şey Avatar. ***Avatar is like a God to ((name_male)).***
10	SM12012	evet tapıyodu o. tapıyodu ona o. bayağı. ***yes, he worshipped it. he worshipped it. a lot.***

By responding with the converging token *aynen*, SM12014 does not attempt to take over the turn, rather he aligns himself with SM12012's evaluations of the issue discussed. By converging, the speaker intends to refrain from any conflictual interaction and maintain good relations with his interlocutor. As illustrated in turn 9, SM12014 shifts the focus of the topic to something else which also affirms that *aynen* is used to mark a topic boundary point.

Engagement function operates on an affective level and reveals the emotive layer of the listener responses. As a result, engagement tokens display addressee's high level of engagement in the speaker's talk. Excerpt (20) below is a mixed group conversation consisting of three 16-year-old speakers, two of whom are male and one of them is female. It is an online interaction, each of the participants reside in different cities in Turkey (Kocaeli, Eskişehir, and Bursa) but attend the same science high school in Eskişehir via distance education due to the COVID-19 pandemic.

(20) Y-3-2M1F-16052021-b

1	SF10015	((name_female))'yla((name_male))'yı karşılaştıramazsın <çünkü aynı kefede değiller. >1> ***you cannot compare ((name_female)) with ((name_male)). <because they are not the same. >1>***
2	SM10006	<((name_female))'yla((name_male))'yı karşılaştırmıyorum ama >1>• şu an ben de • ((name_female))'yla konuşmamız bittiğinde • aynı psikolojik durumdaydım. ***<I am not comparing ((name_female)) with ((name_male)) but >1> I was in the same psychological state, when ((name_female)) and I finished talking.***
3	SF10015	<ya bak. ben X kişisiyle sıradan biriyle konuşsaydım ve biz ilişki... >2> ***<well look. if I were to spoke with a person X and we…>2***
4	SM10005	<yani. bi insanın duygularını belki bir haftada yaşayabilir.>2> <belki üç ayda yaşayabilir. >3> ***<I mean. a person may experience those emotions withi a a week>2> <or maybe in three months time. >3>***
5	SM10006	<ama o benim için>3> sıradan değildi. tamam mı? ***<but to me, she was>3> not ordinary. okay?***
6	SF10015	bi beni dinler misin? ***will you listen to me for a second?***
7	SM10006	psikolojik olarak bi fark yok. ***there is no difference psychologically.***
8	SM10005	ama bişe söylicem. ***but I wanna say something.***
9	SM10006	tamam. benim için sıradan olmadığı için. ben de senin sevgilinden ayrıldığın kadar üzüldüm. <anladın mı? ***okay. because she was not ordinary for me. I felt as sad as you did when you broke up with your boyfriend.<get it? >4>***
10	SM10005	<aynen. >4> ***<exactly.>4>***
11	SM10006	çünkü o benim için sıradan değildi. senin için şu an ((name_male)) neyse • o da benim için oydu. ***because she was not ordinary to me. what ((name_male)) means to you right now, she was to me.***

The topic of the conversation is romantic relationships. When SF10015 shares with her friends that she broke up with her boyfriend, SM10006 consoles SF10015 by stating that

he also has an experience of a heartbreak in turn 2. SF10015 suggests that SM10006 cannot empathize with her as his relationship had lasted shorter than hers did in turn 3. This disagreement receives a resistance from SM10006 who insists that his experience is valid enough to let him empathize with her. In the meanwhile, the other male speaker, SM10005, takes sides with SM10006's arguments in turn 4.

While SM10006 continues defending his argument, SM10005 aims to support him. When SM10006 explains how he felt when he broke up with his girlfriend, SM10005 uses response token *aynen* as an engagement marker to expresses his sympathy with his friend's feelings in turn 10. His non-turn claiming response overlaps (the overlap in this case is indicated by <*I felt sad.* >4> in turn 9 and <*exactly!*>4> in turn 10 in the excerpt) with SM10006's utterance.

A small portion of the response token *aynen* (13%) function as the markers of ***information receipt***. Information receipt tokens are used by speakers at points where they assume that their interlocutor received the message or that the details about the content are understood. The following conversation in (21) exemplifies uses of *aynen* with this function. It is an online interaction between three 18-year-old classmates from an Anatolian high school in Mersin. Speakers are talking about the COVID-19 dashboard which shows the information about daily cases for infections and death in the country.

(21) Y-3-M-06122020-2

1	SM12012	kanka geriden geliyodur bu. ***kanka, it is probably counted retrospectively.***
2	SM12013	aynen◡ .aynen. geriden gelip toplanıyodur. ***exactly exactly. retrospectively and it is cumulatively counted.***
3	SM12012	aynen. ***exactly.***

As presented in the excerpt (21) above, when the speakers check the dashboard, they couldn't figure out how the number of cases are calculated. SM12012 offers an explanation in turn1 to which SM12013 asnwers with convergence token *aynen* in reduplicated form and paraphrases SM12012's explanation to show that he comprehended the way calculation works. In the following turn, SM12012 uses *aynen* as an information receipt token to confirm that SM12013 understood his earlier account of explanation.

Up until this point, the corpus analysis offered an account of the types and distribution of identified response tokens in the CoTY. Formal characteristics of lexical and non-lexical response tokens were presented and exemplified with corpus data. The pragmatic functions of response tokens were discussed with a focus on a register-specific token *aynen* 'exactly' which illustrated that *aynen* displays various layers of pragmatic functions in spoken interaction among Turkish speaking youth. The most frequently used function is that of continuer which suggests that the speakers use it extensively to continue and expand the narrative rather than marking topic boundaries or conveying emotive/commentary signals oriented to speaker's messages. *Aynen* appears in reduplicated form of *aynen aynen* in the data and 38% of reduplications act as continuers which also affirms its prominent function of facilitating the ongoing turn of the speaker. The results also showed that continuer tokens have the supportive function in the co-construction of interactional humour and they are also oriented around face concerns when they are used as convergence tokens in youth language. In the following section, the second group of interactional markers which are *vocatives* will be presented.

4.3.2 Vocatives

Dynamism of spoken interaction is reflected on how youth expresses various interactional labels for each other even though they are friends. Though in earlier works it was suggested that vocatives are not used among "close associates where neither addressee-identifying role nor their relationship-maintanence role is felt to be necessary" (Biber et al., 1999, p. 1112), recent socio-pragmatic as well as corpus-oriented studies exhibit that speakers make use of various addressing practices even in informal and intimate registers of speech. Youth talk proves a rich resource in terms of the range and functions of vocatives and in this line, the following section will outline the scope of vocatives focused in this study, relevant body of research and the results along with extracts from the corpus.

4.3.2.1 Defining vocatives

In dyadic and multi-party spoken discourse, the speakers heavily make use of various linguistic devices to denote their interlocutors to whom the message is addressed. From a sociolinguistic perspective, these linguistic devices contribute to our understanding of co-construction of interpersonal relationships, power hierarchies, (im)politeness,

conveying emotions in discourse. Multiple labels, i.e. 'forms of address' (Brown & Gilman, 1960), 'terms of address' (Ervin-Tripp, 1971), 'address terms' (Jefferson, 1973), 'vocatives' (Leech, 1999), have been used interchangibly to refer to these linguistic devices. Though these labels are closely related, Leech (1999, p. 107) differentiates them syntactically and underlines that a form or term of address is "a device to refer to the addressee(s) of an utterance" while a vocative is a kind of address term which is a "nominal constituent loosely integrated with the rest of the utterance". In other words, vocatives are not syntactically embedded into the argument structure of a sentence (Levinson, 1983, p. 71), thus they are optional elements within a sentence. They can appear in initial, medial, final and stand-alone positions (Clayman, 2012; Leech, 1999) in any type of sentence from declerative to interrogative and imperative (Heyd, 2014).

Vocatives can take form of pronominals (e.g., 2nd person singular pronouns *you* in English and *sen* in Turkish), personal names, and nominal forms. With regard to semantic classification of vocatives, Biber et al.'s (1999) corpus-based study on British and American English offers a comprehensive baseline for different forms of vocatives. In a scale from the most familiar/intimate to the most distant/respectful relationship between the speakers, Biber et al. (1999, p. 1108–1109) categorizes vocatives into endearments (e.g., baby, [my] darling, honey); family terms (e.g., mummy, dad, grandpa); familiarizers (e.g., guys, man, dude, mate, folks, bro); familiarized first names (e.g., Jenny for Jennifer, Tom for Thomas); first names in full (e.g., Jennifer, Thomas); title and surname (e.g., Mr. Graham, Ms. Morrisey); honorifics (e.g., sir, madam); and a category labelled as others which includes nicknames (e.g., boy, lazy).

In terms of its pragmatic functions, Leech (1999) underlines three main pragmatic functions for vocatives, namely (i) summoning attention, (ii) addressee identification, and (iii) establishing and maintaining social relationships between the addresser and the addressee. Elaborating on Leech's (1999) function catagories, McCarthy and O'Keeffe (2003) propose functions of vocatives under two broad levels, namely *organizational* level and *interpersonal* level. Organizational use consists of (i) turn management, (ii) topic management, (iii) summons, (iv) call management; while interpersonal use has (i) badinage (i.e., humourous/witty remarks), (ii) mitigators, (iii) relational as functions. Though this categorization included genre-specific functions and was based on English from Britain and Ireland, taxonomy was also utilized to investigate vocatives in other

languages and contexts (Clancy, 2015 on Irish English; Palacios Martínez, 2018, 2021 on Spanish; Tsakona & Sifinaou, 2019 on Greek).

This study focuses on nominal type of vocatives in youth language and presents the structural, semantic and functional characteristics of these vocatives in Turkish youth language. Following a brief overview of recent studies on vocatives, the results will be presented and exemplified within this scope.

4.3.2.2 Brief overview of related work on vocatives

The overview of vocative studies will be briefly presented in two sub-sections: recent work on studies on Turkish vocatives and youth language research focusing on vocatives.

4.3.2.2.1 Vocatives in Turkish

Though few in number, studies on Turkish vocatives, which use the terms 'vocative' and 'address term' interchangeably, have deployed a wide range of tools and approaches to investigate these linguistic devices so far. The studies so far have proposed systems to describe the pronominal addressing system in Turkish (Yüceol Özezen, 2019), focused on factors influencing the choice of T/V forms such as age, gender, group membership (Balpınar, 1996; König, 1990) and explored pragmatic functions of vocatives such as showing emotions and solidarity (Hatipoğlu, 2008; Yüceol Özezen, 2019).

The earlier studies heavily made use of elicited data sources. Bayyurt and Bayraktaroğlu (2001), for instance, examined pronouns and other address terms in service encounters in Turkish using discourse completion tasks, questionnaire data, and field notes. The researchers noted gender differences in T/V uses as female customers favoured V pronoun in more contexts compared to males who overall preferred T pronoun. The interactional goal, which is completing a transaction, was also argued to have an effect on the T/V forms used. The study underlined dimensions of power and solidarity as the main determining factors for choosing an address term. Similarly, Zeyrek (2001) took an emic perspective on T/V forms of address in Turkish and argued that *sen* (T) is used to encode solidarity while *siz* (V) indicates asymmetrical power relationship between the speakers. Based on questionnaire responses collected from undergraduate students, Aktaş and Yılmaz (2016) list and exemplify the address terms reported. Though the

researchers adopt a prescriptive approach to the address terms used, their results echo Özbay and İpek's (2015) observations and provide exploratory examples of contemporary uses of address terms such as *kanka*. Dimension of creativity in vocatives was explored in Çetintaş Yıldırım's (2018) work on child-directed speech used by parents. The results of interview data suggested that parents generate various creative variants of conventional address terms while addressing their children. The study also notes that negative address terms are used to show intimacy. In a similar vein, Gökşen's (2015) work discusses addressing practices in various provinces in Anatolia. The study argues that men use negative address terms such as *kül dökücü* 'ash shedder', *çorbacı* 'soup maker', *kanayaklı* 'bloodfoot' for their wives to 'protect them from any harm' which is discussed within the scope of patriarchy and discourses of masculinity. The dynamic nature of vocatives was reported by Alkan Ataman (2018) in her doctoral dissertation which adopted a comparative approach to classify address terms in Turkish. Combining Old Turkish, Old Oghuz Turkish, Ottoman Turkish, and contemporary Turkish data sources, she highlights the evolving nature of addressing practices over time. The study also mentions unconventional uses of address terms and classifies the divergences under semantic, structural and cultural categories.

Studies which make use of naturally occurring data complement the aforementioned accounts of research on vocatives. Research which captures the interactional facet of language reflect the dynamic nature of vocatives. Within this scope, Alaca's (2014) exploratory MA thesis investigated address terms used in a specific genre of television reality shows in Turkey, namely matchmaking programmes. The study indicates that speakers can switch between fictive kinship terms and honorifics based on the context and age and the level of education. In Özcan's (2016) longitudinal study in which naturally occurring data from 56 monolingual Turkish and 48 monolingual Danish speaking students are analysed, functions of several address terms are listed. The study states that first names are used for getting attention, giving instruction, warning and requesting while diminutives are used to show affection. The address terms were also classified into positive, negative, neutral clusters in which positive address terms were identified to appreciate good behaviour and negative address terms were used for criticism. The lexical items *ulan* and *kız* are classified as netural address terms which showed wide range of functions such as expressing emotions, stating (dis-)agreement, persuading and demanding. Using corpus data from both the STC and the TNC, Işık-Güler and Eröz-Tuğa (2017) provided a comprehensive account of *ulan* and described it and

its variants within the scope of (im)politeness theories. Among various interactional functions of *ulan* are listed the vocative use and indexing relationship between speakers. As another corpus-based contribution to the study of vocatives, Özer (2019) explored 9.5 million-word sub-corpus of imaginative prose and 1 million-word sub-corpus of spoken language from the TNC and classified all types of address terms and their functions identified in the corpus. The study lists the functions of Turkish address terms as involving agreement, attention gathering, conveying the feeling, holding the floor/foregrounding the talk, involving non-conforming utterances/disagreements, making the listener remain focused, selecting next speaker, situational role designation/setting the tone of the communication, softening the virtual commands, and topic shifting. The study is prominent in the sense that it offers a comprehensive account for the study of address terms for Turkish language within the timeframe of 1990 to 2009.

Vocatives are inherently interactional and the overview of studies in Turkish indicate that there seems a recent inclination of utilizing naturally occurring data to explore vocatives. Yet research on interactional spoken data is still scarce and thus needs further scholarly investigation.

4.3.2.2.2 Vocatives in youth talk

Studies on youth language frequently mention that vocatives are among the prominent linguistic features identified in the interaction between young speakers. In this line, patterns of vocative uses in English, Spanish and partly Norwegian have been extensively studied thanks to the available spoken corpora of young speakers of these languages. These studies have investigated variables of gender, socioeconomic background, and language in their comparative studies of vocatives in youth talk. The COLT and the COLA data indicated that both male and female teenagers from Madrid use more vocatives than those in London and that speakers use these vocatives as intimacy markers and to maintain social contact among their groups (Jørgensen, 2010, 2013; Rodríguez-González & Stenström, 2011; Stenström & Jørgensen, 2008).

Among types of vocatives, taboo vocatives have gathered particular attention in youth talk. Hasund & Stenström (1997) and Stenström et al. (2002) explored the use of vocatives in ritual conflict exchanges between female teenagers in the COLT and

presented the parameter of socioeconomic background as a one which influences the use of taboo vocatives. Though limited in terms of speakers (*n*=4), the study points out at the working-class girls used sexual abuse words when addressing their interlocutors while middle-class girls did not in their speech.

Palacios Martínez's (2011a) work on the COLT data also show that abuse and insult words are frequently used as vocatives among English speaking teenagers with the goal of enhancing solidarity. As a pattern in English, teenagers often used taboo vocatives preceded by the pronoun *you* and even though the COLT did not provide data for the use of taboo vocatives in mix-gender interactions, it was reported that boys used these vocatives more frequently than girls did. In a subsequent work, Palacios Martínez (2021) adopted a more comprehensive approach to investigate taboo vocatives and identified a total of 59 types in the COLT. Among the identified vocatives, the majority of them had sexual references or were related to sexual behaviour. In terms of their pragmatic functions, the study underlines that taboo vocatives are not always used as straightforward insults but rather they are used to organize discourse, express contempt and envy, reinforce affection and badinage, and enhance in-groupness.

In addition to corpora-based studies, Günther's (2011) work which collected data from informal interactions between male youth with migrant backgrounds also mentions that young male speakers of German used 'insulting remarks' when they address each other. These lexical items were frequently observed in greetings and usually accompanied laughter which suggested that these practices are closely related to the social and cultural identities manifested in transmigrational contexts. In this particular study, the researcher examines an excerpt in which a speaker from Turkish background is reported to perform an insult ritual with the purpose of negotiating hierarchy in a group. This observation regarding using insult terms as vocatives was also previously argued to be a conventionalized cultural form of displaying masculinity, dominance, coolness for Turkish male youth (Dundes, Leach, & Özkok, 1972).

There are also studies which specifically collected data from speakers who are labelled as 'young speakers' or 'youth' with the aim of focusing on particular vocatives. Among these, Kiesling (2004) outlines the patterns of *dude* in American English with a particular focus on practices of young male speakers. The study indicates that *dude* is used as a stance marker and indexes a stance what the researcher terms as 'cool solidarity'

constructed around discourses of young masculinity. Based on the analysis of field notes of everyday talk and self-report surveys in the University of Pittsburg, Kiesling (2004) states that there is a dominance of male-male uses of *dude* and it is less frequently used by females and to females. With regard to construction of 'cool solidarity', the functions of *dude* are listed as marking discourse structure, exclamation, confrontational stance mitigation, marking affiliation and connection and signalling agreement. The study argues that *dude* encodes masculinity in North America among youth. Heyd (2014) carried out an exploratory cross-linguistic investigation of vocatives in contemporary slang and compared *dude* in American English and *alter* in German. The researcher made use of a composite set of data (corpus samples, previously published data, online discourse data, meta-communicative data) and corroborated the arguments of Kiesling (2004) in terms of function of *dude* with regard to conveying 'cool solidarity' as the results demonstrated that *dude* was indexical for certain youth groups and implied a sense of in-groupness. Another semantically close vocative is *mate* which has been extensively investigated in Australian English by studies of Rendle-Short (2009, 2010). Though traditionally conceptualized within the discourses of male comradeship, *mate* as an address term was reported to display a shift in its conceptualization and use in Australia. It was reported that younger generation of women (ages 18-29) used this particular address term while those who were 50 years and above did not. Based on the self-reported functions, young women used *mate* to address their interlocutors of both genders and they associated the term with friendliness, fun and intimacy. Additionally, it was also underlined that *mate* acts as a marker of Australian in-group identity by speakers of other languages. Parkinson's (2020) study also contributes to research on *mate* from a focus on a different interactional context, namely classroom discourse in New Zealand. The study investigated *mate* and *guys* and concluded that these vocatives had different pedagogical functions in interaction. Vocative *guys* had instructional functions in teaching such as attracting attention, marking boundaries and transitions in a task, highlighting important content while *mate* was utilized for mitigating and affective functions such as making criticism, giving praise and encouragement.

Altogether, the growing body of literature on vocatives in youth talk draws attention to various intertwined factors behind the selection of vocatives with a growing focus on naturally occurring spoken data. While current studies which mainly concentrate on English and Spanish provide an outline for the description and uses of vocatives in youth interaction, a wider perspective covering other languages is required to compare and confirm the findings of the existing body of literature.

4.3.2.3 Findings: Vocatives in the CoTY

In this section, the identified nominal vocatives in the CoTY will be outlined. The types, addressers, addressees, forms, positions, reciprocity, semantic categories of these vocatives will be presented. As the presented features of vocatives show high degree of intersectionality, the final part of this chapter explores the role of vocatives in achieving interactional goals among young speakers of Turkish by focusing on the most frequently occurring vocative in the corpus kanka 'dude'. Additionally, the phenomenon of address shifts, displaying shifts of address for the same addressee in conversation, was identified in the CoTY. These shifts will be exemplified and discussed for the case of vocatives bro (borrowing from English) compared to kardeş 'sibling' which are semantically equivalent lexical items but both are in use in Turkish youth language.

4.3.2.3.1 Types, addressers, and addressees

In the CoTY, a total of 48 types of 2111 tokens of vocatives were identified. Keywords to identify vocatives were generated based on both the existing literature (See 4.3.2.2) and emergent findings listed during data transcription and annotation process. For each keyword, corpus queries were run and each concordance line was manually examined to determine whether the target lexical item was used as a nominal vocative or not (see 4.3.2.1). Those which were not categorized as vocatives were excluded from the list. In Table 24 below, total number of tokens retrieved from the corpus is presented (TN) and the number of identified vocatives is presented with their absolute frequencies (AF) along with the relative frequencies (RF) per million. For each type of vocative, total number of unique speakers, number of female speakers and male speakers are also presented to illustrate the extent each vocative is used by the speakers of the CoTY.

Table 24 Distribution of vocatives in the CoTY

Rank	*Vocative (type)*	*Gloss*	*No. of tokens*				*No. of addressers*					
			All occurences		*Vocatives*		*All*		*Female*		*Male*	
			TN	*RF*	*AF*	*RF*	*N*	*%*	*N*	*%*	*N*	*%*
1	**kank(a/i/a/s)**	*dude*	702	4160.05	**680**	4029.67	68	55	30	48	38	62
2	**oğlum**	*my son*	454	2690.40	**452**	2678.55	58	47	10	16	48	79
3	**abi(m)**	*(my) elder brother*	399	2364.47	**302**	1789.65	59	48	26	42	33	54
4	**(u)la(n)**	*deriv. boy*	329	1949.65	**299**	1771.87	65	53	17	27	48	79
5	**aslanım**	*my lion*	84	497.78	**84**	497.78	3	2	0	0	3	5
6	**kardeş(im)**	*(my) brother*	110	651.86	**58**	343.71	26	21	6	10	20	33
7	**bro**	*bro*	38	225.19	**38**	225.19	5	4	2	3	3	5
8	**arkadaş(ım/lar)**	*(my) friend(s)*	228	1351.13	**27**	160.00	16	13	6	10	10	16
9	**canım**	*my dear*	58	343.71	**26**	154.08	18	15	10	16	8	13
10	**aga**	*deriv. elder brother*	20	118.52	**20**	118.52	11	9	3	5	8	13
11	**salak**	*stupid*	79	468.15	**16**	94.82	10	8	5	8	5	8
12	**reis**	*chief*	14	82.96	**13**	77.04	1	1	0	0	1	2
13	**birader**	*brother*	9	53.33	**8**	47.41	3	2	0	0	3	5
14	**bebeğim/bebişko**	*(my) baby*	8	47.41	**7**	41.48	5	4	2	3	3	5
15	**gerizekalı**	*idiot*	56	331.86	**6**	35.56	5	4	3	5	2	3
16	**hacı/hacıt**	*pilgrim*	13	77.04	**6**	35.56	4	3	0	0	4	7
17	**beyler**	*gentlemen*	5	29.63	**5**	29.63	4	3	0	0	4	7
18	**aşkım/aşko**	*(my) love*	4	23.70	**4**	23.70	3	2	2	3	1	2
19	**kız(ım)**	*(my) girl*	133	788.16	**4**	23.70	2	2	2	3	0	0
20	**pezevenk**	*pimp*	8	47.41	**4**	23.70	3	2	1	2	2	3
21	**baba**	*father*	260	1540.76	**3**	17.78	3	2	0	0	3	5
22	**gençler**	*guys*	15	88.89	**3**	17.78	3	2	1	2	2	3
23	**koçum**	*my ram*	4	23.70	**3**	17.78	3	2	1	2	2	3
24	**mal**	*dumb*	43	254.82	**3**	17.78	3	2	0	0	3	5

Table 24 (cont'd)

Rank	Vocative (type)	Gloss	No. of tokens				No. of addressers					
			All occurences		Vocatives		All		Female		Male	
			TN	RF	AF	RF	N	%	N	%	N	%
25	**moruk**	*geezer*	3	17.78	**3**	17.78	1	1	0	0	1	2
26	**puşt**	*prick*	6	35.56	**3**	17.78	2	2	0	0	1	2
27	**sikik**	*fucked*	3	17.78	**3**	17.78	3	2	0	0	3	5
28	**şerefsiz**	*undignified person*	10	59.26	**3**	17.78	3	2	0	0	3	5
29	**yavrum**	*my baby*	3	17.78	**3**	17.78	3	2	0	0	3	5
30	**anam**	*my mother*	6	35.56	**2**	11.85	2	2	2	3	0	0
31	**çocuğum**	*my child*	5	29.63	**2**	11.85	1	1	1	2	0	0
32	**kerata**	*rascal*	3	17.78	**2**	11.85	2	2	1	2	1	2
33	**manyak**	*crazy*	29	171.85	**2**	11.85	2	2	1	2	1	2
34	**usta**	*master*	16	94.82	**2**	11.85	2	2	0	0	2	3
35	**yavşak**	*imposing person*	8	47.41	**2**	11.85	2	2	0	0	2	3
36	**dostum**	*my friend*	3	17.78	**1**	5.93	1	1	0	0	1	2
37	**evladım**	*my child*	1	5.93	**1**	5.93	1	1	0	0	1	2
38	**ezikler**	*losers*	1	5.93	**1**	5.93	1	1	0	0	1	2
39	**göt**	*ass*	4	23.70	**1**	5.93	1	1	0	0	1	2
40	**güzelim**	*my beauty*	2	11.85	**1**	5.93	1	1	1	2	0	0
41	**hatun**	*woman, wife*	1	5.93	**1**	5.93	1	1	0	0	1	2
42	**hayatım**	*my life*	49	290.37	**1**	5.93	1	1	1	2	0	0
43	**hocam**	*my teacher*	54	320	**1**	5.93	1	1	0	0	1	2
44	**ibne**	*fagot*	5	29.63	**1**	5.93	1	1	0	0	1	2
45	**kuzu**	*lamb*	3	17.78	**1**	5.93	1	1	1	2	0	0
46	**millet**	*guys*	29	171.85	**1**	5.93	1	1	0	0	1	2
47	**şekerim**	*sweetie*	1	5.93	**1**	5.93	1	1	1	2	0	0
48	**tatlım**	*sweetie*	1	5.93	**1**	5,93	1	1	1	2	0	0
	Total		*3319*	*19668.38*	***2111***	*12515.70*	*123*	*100*	*62*	*100*	*61*	*100*

TN: Total number of tokens in corpus, *AF:* Absolute frequencies of vocatives, *RF*: Relative frequency per million

The results show that most frequently occurring vocatives in Turkish youth language are *kanka* and its variants *kanki, kank, kanks* (*AF*=680, *RF*=4029.67)*;* followed by *oğlum* (*AF*=452, *RF*=1789.65); and abi (*AF* =302, *RF*=1789.65). In the CoTY, 19 out of identified 48 vocatives (40%) are present in the speech repertoires of both female and male speakers. The vocatives which are used by highest number of speakers in the corpus are *kanka* and its variants (*n*=68); *ulan* and its variants (*n*=65); and *abi* (*n*=59). Vocative *kanka* is used by the 48% of all female speakers in the CoTY which makes it stand out as the vocative type used by the young female speakers of Turkish most extensively. Among male speakers, *oğlum* along with *ulan*[32] and its derivatives *lan* and *la* are the most extensively used vocative types as 79% of all male speakers in the corpus use these vocatives.

4.3.2.3.2 Forms and positions

As also indicated in Table 24, some vocatives in the CoTY are marked with two kinds of inflectional morphemes: possessive and plural suffixes. The results show that a group of nominals (*aslan, ana, aşk, bebek, can, çocuk, dost, evlat, güzel, hayat, hoca, koç, oğul, şeker, tatlı, yavru*) only occur with the inflectional morpheme of 1st person singular possessive marker *-(I)m* while another group *(abi, arkadaş, kanki, kardeş, kız)* occur both in their base forms and with possessive suffixes for 1st person singular forms in data. This nominal inflectional suffix indicates the possessor of the marked noun, as illustrated in (i) below:

(i) kardeş -im
brother -1SG.POSS POSSESSION
'my brother'

Second inflectional morpheme is the plural suffix *-lAr* which marks number in Turkish. Though low in number, a group of nominals (*bey, genç, ezik*) are used only with plural suffixes in the CoTY while nominal *arkadaş* does not necessarily require plural marking for it to be used as a vocative. It should also be noted that there are vocatives (*lan, millet*) which are not marked with a plural suffix but are used to refer to groups of addressees. An example of plural suffixation is presented in (ii) below:

[32]a lexical item reported to be etymologically derived from *oğlan* (Eng. boy) and marked as a part of contemporary Turkish slang (see Nişanyan, 2012).

(ii) bey -ler
gentleman -PL
'gentlemen'

Due to the fact that the CoTY consists of spoken data which displays constant co-construction of interaction, vocatives are not frequently found as isolated utterances (only 2%) in the corpus data, but rather they occur within the proximity of utterances, namely utterance initial, medial and final positions, which also suggests that they have supportive roles in discourse (McCarthy & O'Keeffe, 2003, p. 159). Echoing the observations reported by the previous studies on corpora data (Biber et al. 1999; Leech 1999; McCarthy & O'Keeffe 2003, Palacios Martínez, 2021), the most frequent utterance position of the vocatives in the CoTY is identified as final position (40%) followed by initial position (33%).

In Table 25 below, the most frequently occurring 10 vocatives and the distribution of utterance positions are presented.

Table 25 Distribution of positions of vocatives in utterances

Rank	*Vocative*	*Stand-alone*		*Initial*		*Medial*		*Final*		*N*
		N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	
1	**kank(a/i/s)**	20	40	398	56.8	79	15.5	183	21.4	680
2	**oğlum**	2	4	211	30.1	70	13.2	169	19.8	452
3	**abi(m)**	0	0	10	1.4	175	34.3	117	13.7	302
4	**(u)la(n)**	0	0	12	1.7	90	17.6	197	23	299
5	**aslanım**	2	4	22	3.1	15	2.9	45	5.2	84
6	**kardeş(im)**	8	16	10	1.4	22	4.3	18	2.1	58
7	**bro**	1	2	8	1.1	16	3.1	13	1.5	38
8	**arkadaş (ım/lar)**	2	4	5	0.7	10	1.9	10	1.1	27
9	**canım**	1	2	0	0	9	1.7	16	1.8	26
10	**aga**	0	0	1	0.1	9	1.7	10	1.1	20
	Others	14	28	23	3.2	13	2.5	75	8.7	148
	Total	***50***	*2*	***700***	*33*	***508***	*24*	***853***	*40*	***2111***

As previously mentioned, not all vocatives occupy the whole turn in data and among them, the most frequently observed stand-alone vocative is *kanka* 'dude' throughout the corpus followed by *kardeş* 'sibling'. *Kanka* also occupies majority of utterance-initial positions followed by *oğlum*. In terms of vocatives with utterance-medial positions, *abi*

'elder brother' is the most frequently observed vocative and in final position, vocative *(u)la(n)* stands out in the corpus.

4.3.2.3.3 Semantic categories

The vocatives in the CoTY fall under five semantic categories. Categories of 'endearment', 'familiarizer', and 'titles' are adapted from Biber et al.'s (1999) list developed for English; the category of 'insult' is generated based on the existing literature on taboo vocatives (Günther, 2011; Hasund, 1997; Palacios Martínez, 2011a, 2021; Stenström et al., 2002) and the term 'fictive kinship' is borrowed from the field of anthropology which was also used by Braun (1988) in her cross-linguistic discussions of address terms. The main categories, sub-categories, and corresponding tokens are presented in Table 26 below. The expanded contexts of the representative instances for each semantic category shows that the vocatives can display a range of pragmatic functions which can also diverge from their semantic categorization.

Table 26 Semantic categories of vocatives

Semantic category	*Sub-category*	*Token*	*Gloss*
familiarizer (*n*=1018)	**single**	arkadaş(ım)	*(my) friend*
		dostum	*my friend*
		kank(i/a/s)	*dude*
		kız	*girl*
		(u)la(n)	*derivative of 'boy'*
	group	beyler	*gentlemen*
		gençler	*guys*
		millet	*friends*
		arkadaşlar	*friends*
fictive kinship (*n*=898)	**offspring**	bebeğim	*my baby*
		çocuğum	*my child*
		evladım	*my child*
		kızım	*my daughter*
		oğlum	*my son*
		yavrum	*my child*
	parents	anam	*my mother*
		baba	*Father*
	sibling	abi(m)	*(my) elder brother*
		aga	*derivative of elder bro.*
		birader	*brother*
		bro	*brother*
		kardeş(im)	*(my) sibling*

Table 26 (cont'd)

Semantic category	*Sub-category*	*Token*	*Gloss*
endearment (*n*=123)	**animal**	aslanım	*my lion*
		kuzu	*lamb*
		koçum	*my ram*
	life	canım	*my dear*
		hayatım	*my life*
	partner	aşkım	*my love*
		hatun	*woman, wife*
	physical	güzelim	*my beauty*
	taste	şekerim	*my sweetie*
		tatlım	*my sweetie*
insult (*n*=50)	**age**	moruk	*geezer*
		kerata[33]	*rascal*
	behaviour	ezikler	*losers*
		gerizekalı	*idiot*
		mal	*dumb*
		manyak	*crazy*
		salak	*stupid*
		şerefsiz	*undignified person*
		yavşak	*impossing person*
	body	göt	*ass*
	sexual	ibne	*fagot*
		pezevenk	*pimp*
		puşt	*prick*
		sikik	*fucked*
title (*n*=22)	**leadership**	reis	*chief*
	occupation	hocam	*my teacher*
		usta	*master*
	religious	hacı	*pilgrim*

The familiarizers are groups with the highest number of vocatives (*n*=1018) which can denote both singular addressees such as *dostum* 'my friend' and group of addressees such as *millet* 'guys'. Familiarizers have been previously defined to be marking acquaintance (Leech, 1999) and solidarity (Palacios Martínez, 2021; Wilson, 2010) as in (22) where speaker uses familiarizer *lan* to address his two male friends and shares his intention to spend time with them. Vocative *lan* here signals the close relationship between the speakers.

[33] The official dictionary for Turkish published by the Turkish Language Association provides the first meaning of this lexical item as '*a man cheated by his wife'* yet the conventionalized meaning of this item is closer to the third entry which defines it as a form of reproaching younger children. Therefore, the token is coded in the age sub-category of the semantic category of insults.

(22) Y-3-M-07102021

1	SM10002	((laughs)) geliyim mi lan yanınıza? çalışmaya değil ama ziyarete. ***((laughs)) shall I come over you, lan? not to work but to visit.***
2	SM10017	((XXX))
3	SM10002	((short laugh)) ben amelelik yapmam! ***((short laugh)) I won't do manual labour!***
4	SM10011	gel oğlum. ye yediğin kadar. sonra gidersin. ***come, man. eat as much as you want. then you can leave.***

Additionally, the results also suggest that they may take other pragmatic roles as in example (23) below in which speaker uses the vocative *dostum* 'my friend' as a mitigator in a face threatening act directed at another speaker in the conversation.

(23) Y-3-M-02122020-c

SM11012	biri şuna Snap öğretsin! telefonun icadından haberi yok. ((laughs)) dostum bu çocuk köylü! ***somebody teach him Snapchat! he does not have any clue about mobile phones. ((laughs)) my friend, this kid is a countryboy!***

Fictive kinship vocatives are family terms which are used for non-relatives. In the CoTY, semantic category with the second highest number of tokens is fictive kinship vocatives (*n*=898) which is also rich in terms of number of types of tokens (*n*=13). Within the sub-category of offspring, vocatives *bebeğim* 'my baby', *çocuğum, evladım, yavrum* (which all correspond to 'my child' in English), *kızım* 'my daughter/girl' and *oğlum* 'my son' are observed. In the sub-category of parenthood, lexical items *anam* 'my mother' and *baba* 'father' are present. Example (24) below exemplifies the use of *baba* as a vocative in male-male interaction. The extract below also provides an episode of address shift in which speaker initially uses the familiarizer *kanka* in turn 1 and then shifts to *baba* in turn 3 to address the same person in his following turn (see section 4.3.2.3.6 for more detail on address shifts)

(24) Y-2-M-05122020

1	SM11008	kanka onu geç. ben sana ne dicem. hani iki bin yirmi bire… ***kanka never mind that. I'll tell you something. you know for two thousand twenty two…***
2	SM11009	hah˙ ***yeah.***

3	SM11008	baba şarkıyı çıkarıyoruz. ***father, we are releasing the song.***
4	SM11009	((laughs)) cebimde iki lira. ***((laughs)) two liras in my pocket.***
5	SM11008	kanka ciddi. bu arada ciddi diyom. ***kanka, seriously. I am serious by the way.***

The sub-category of sibling covers the vocatives *abi* 'elder brother' and *birader* 'brother', *kardeş* 'sibling' and *aga* (old Turkish for elder brother, currently part of contemporary slang). There is also English vocative *bro* which is used by and for both genders in the corpus even though Turkish equivalents for *bro,* i.e., *kardeş* and *birader* are also used as vocatives in the corpus. It is interesting to note that the speakers who use vocative *bro* (*n*=5, 2 female and 3 male speakers) also use the vocatives *abi*, *aga, kardeş(im)* for the same interlocutors. Though limited in terms of number, these shifts highlight the situatedness of vocatives in interaction of youth and lay basis for further analysis of their pragmatic functions (see section 4.3.2.3.6 for pragmatic functions of *bro* and *kardeş*).

As for endearment vocatives (*n*=123), the results show that speakers make use of the source concepts of *animals, partner, taste, physical appearance* and *self* to convey positive feelings towards their interlocutors, thus they have positive semantic prosody in the corpus. Animal endearment terms highlight specific characteristics of animals as in cases of vocatives *aslan* 'lion' and *koç* 'ram' which are associated with power in Turkish culture (Alkan Ataman, 2018) and *kuzu* 'lamb' for features of cuteness[34]. In contrast with what was observed for Serbian (Halupka-Rešetar & Radić, 2003), animal names as vocatives are not derogatory in interaction between Turkish speaking youth as in example (25) below. In a conversation between a male and female speaker who talk about national university entrance exam and the rankings for faculties of medicine in Istanbul, vocative *koçum* is used by the female speaker to give reassurance to and boost confidence of her male friend.

(25) Y-2-FM-04122020

[34] Though not identified in the CoTY data, this study recognizes other forms of animal vocatives frequently used in Turkish such as *köpek* 'dog', *ayı* 'bear', *yılan* 'snake', *inek* 'cow', *kuş* 'bird', among many others. An example is from the written Turkish data of TNC: *Zavallı kumral kuşum benim* 'My poor brunette bird' in which the author uses an animal name to addresses a girl in a fictional narrative [W-KA16B0A-0118-103].

SM10004 İstanbul'da gerçi böyle • üç dört bin • beş bin falan tıp var mı ki?
in Istanbul, though, are there such (faculty of) medicine like three, four thousand, five thousand (range)?

SF11006 vardır ya! illa. hemen bakarız koçum. ‿ayıp ediyorsun ((chuckles))
there must be! let's check, koçum. no worries ((chuckes))

Within the scope of endearment vocatives, a culturally marked lexical item, *canım* is also observed as an endearment vocative among friends. The vocative corresponds to *my dear* in English but it is literally translated as 'my heart' which is related to the concept of 'essense of self' in Turkish culture. Existing corpus-based work illustrates that this lexical item displays relational functions of emphasizing agreement, mitigating face threats, and converging with the interlocutor (Efeoğlu, 2019). Similarly, the use of *canım* in the CoTY appears to have diverged from this primary highly affection-laden meaning and carries additional context-dependent functions such as in (26) where speaker uses it within a response to a face-threatening act (FTA). In the conversation below, SF09003 scolds SF09004 for posting a visual with sensitive content on social media. SF09004 accepts the FTA and the vocative *canım* has a supportive role as an attempt to maintain the harmonious relationship with her interlocutor.

(26) Y-2-F-02122020

SF09003 hı-hı˙ e yani sen de salak mısın. ‿koyuyosun?
mm-hmm˙ and are you such an idiot to put it there?

SF09004 sorman kabahat • canım.
you don't need to ask, my dear.

It should be noted that all the fictive kinship vocatives except for *baba* 'father' and all the endearment vocatives except for *kuzu* 'lamb' take possessive suffixation *–(I)m* in corpus. As was previously reported with regard to use of possessive determiners with vocatives in Spanish (Fernanádez-Mallat, 2020, p. 98), this suffixation can be associated with conveying high levels of affection in Turkish as well. By integrating possessive suffixation for fictive kinship and endearment terms, the conveyed feelings of affection and intimacy are intensified.

Confirming the recent work on vocatives in youth language, the results show that the use of insult vocatives are not gender-exclusive as both females and males use them as vocatives in interaction. Among users of insult vocatives, 15 speakers are male and 10

speakers are female and these vocatives are manifested in both same-sex and mixed interactions.

Semantic category of insults (*n*=50) covers lexical items which refer to humiliations with regard to age such as *moruk* 'geezer', types of behaviour which refers to either lunacy such as *gerizekalı* 'idiot' or *şerefsiz* (lit. undignified person); as well as taboo words such as *göt* 'ass', and sexually connotated references to people such as *pezevenk* 'pimp' and *sikik* 'fucked'. This categorization also illustrates that insults are also the richest vocative category in terms of number of types of tokens (*n*=14), the variety of insult types identified are more diverse than other semantic categories. Among other functions, the vocatives in this category can be used with a humiliating function as in (27) where speakers are engaged in a conflictual talk and SF09006 female speaker uses the vocative *pezevenk* when she responds to an utterance of verbal aggression.

(27) Y-2-F-21112019

1	SF09005	tamam. no problem. sıkıntı yok. ***okay. no problem. no problem.***
2	SF09006	no problem diyosun ama benim sinir katsayılarımı artırıyosun. sonra kavga ediyoruz! ***you say it's not a problem but you are getting on my nerves. then we start fighting!***
3	SF09005	tamam boş yapma! ***okay, just cut it out!***
4	SF09006	sen boş yapma asıl pezevenk! gerizekalı! ***you cut it out, pimp! moron!***

Additionally, insult vocatives are also used in humourous interaction among Turkish speaking youth. An example is (28) where SM12006 asks his friend the reason why he never replies back to his messages. His friend SM12007 explains that he is studying intensively so he does not notice the messages. SM12006 does not accept this as a valid excuse and teases his friend integrating the vocative *pezevenk* to express his disagreement with his interlocutor with a humourous tone. The expanded context following the turns include laughter and endearment terms which also confirm that the vocative did not convey an offensive tone in interaction.

(28) Y-3-M-26112020

SM12007 sen senin mesajını mı diyon? ((XXX)) çalışıyom oğlum! tabii senin mesajını mı görecem?
are you talking about the messages you sent me? ((XXX)) I am studying, man! how would I notice your messages?

SM12006 kızdan mesaj gelirse hemen görüyon pezevenk!
you notice the messages if they are from the girls, you pimp!

Additionally, in line with the previous research which mentions that socioeconomic status could be an influencing factor in the use of taboo vocatives (Hasund & Stenström, 1997), insult vocatives in the CoTY also shows a tendency to be present in speaker groups with lower socioeconomic backgrounds than those in higher socioeconomic backgrounds (*n*=32 and *n*=6, respectively).

The final semantic category is titles (*n*=22) which refers to nominals traditionally used to mark the occupation or rank of a referent. In the Turkish context, these vocatives are also reported for denoting elder speakers without necessarily indexing any occupation (Alkan Ataman, 2018). In the CoTY as well, these vocatives do not refer to actual titles of the speakers but rather they are desemanticised and act as pseudo titles among friends. This group of vocatives were exclusively observed in male-male interaction. In (29) below, *hacıt,* a variant slang form of vocative *hacı* 'pilgrim', is used by a speaker to hold the floor while conveying his argument to his interlocutor.

(29) Y-2-M-06122020-b

1 SM12001 çok ha! diş hekimliği ne kadar?
that is too long! how long does dentistry take (undergraduate programme)?

2 SM12002 hiç bilmiyom ki. dörttü herhalde. genelde dört oluyo. ben bi tıpı biliyorum. altı sene. ((3.0)) o kadar.
I don't know at all. I guess it is four (years). it is usually four. I have some knowledge only on the medicine. it is six years. ((3.0)) that's all.

3 SM12001 tıp da çok ya! hacıt. eşit ağırlık devam. ((short laugh))
medicine is too long! hacıt. let's stick with our own track. ((short laugh))

All in all, initial observations within each semantic category of vocatives suggest that the pragmatic functions of vocatives are multifaceted and context dependent in Turkish youth language. While it is possible to track the fundamental semantic associations of

some vocatives such as (25) in interaction, there are instances where vocatives undergo desemanticization in interaction.

4.3.2.3.4 Referents

The results also show that even though some vocatives are semantically marked for gender in Turkish, they are used by speakers to address both female and male interlocutors in interaction. To elaborate, *abi* 'elder brother' and *oğlum* 'my son' both have originally gendered (masculine) referents in Turkish. As vocatives in the CoTY, they are extensively used (*n*=59, *n*=58, respectively) to denote both female and male referents. This observation suggests a similar pattern previously noted for *man* by Cheshire (2013) in the sense that pronouns, in this case vocatives, undergo pragmatic extension by losing their gender referent and as a result are more widely used for both males and females. In Cheshire's study, it is suggested that this desemanticisation leads to the intertwined functions of *man* as an address term and a pragmatic marker. Similarly for Turkish data, the analysis suggests pragmatic extension of vocatives considering the referents of originally masculine vocatives in the corpus. In the CoTY, vocatives which have masculine referents[35] (*abi, aga* 'elder brother'*, baba* 'father'*, beyler* 'gentlemen'*, birader* 'sibling'*, bro, oğlum* 'my son', *pezevenk* 'pimp') are used by speakers from both sexes for both male and female referents. For instance among these vocatives, *abi* 'elder brother', which is used by a relatively balanced number of speakers in the corpus (26 female and 33 male speakers), has 27 female and 35 male addressees in the corpus (See Appendix I for the distribution of addressee and addressers for aforementioned vocatives). Vocatives which originally had feminine addressees (*kız* 'girl'*, anam* 'my mother'*, hatun* 'woman') do not display this pattern yet it should be underlined that the scope of data is limited in terms of the number of tokens for this cluster of vocatives (*n*=7) compared to number of tokens for vocatives with masculine referents (*n*=841).

That being said, the observations regarding pragmatic extension observed for vocatives with male referents suggest that gender of the vocative type, sex of addresser and the addressee are not determining factors in selection of the type of vocative to be used in informal conversation among friends in Turkish, rather there is an interplay of context and interactional goals which shape the patterns for the uses of vocatives.

[35] Vocatives *ibne* 'fagot', *kerata* 'rascal', and *puşt* 'prick' are excluded from the analysis as they are identified to be idiosyncratic uses.

4.3.2.3.5 Multiplicity of functions: *kanka*

The initial observations regarding semantic categories and the instances of pragmatic extension indicated that vocatives are manifested in multi-faceted interactional space in Turkish youth language. To explore the situatedness and identify the pragmatic patterns of vocatives, the most frequently occurring vocative *kanka* 'dude' and its variants, which accounts for 32% of all vocatives in the corpus, were analysed. It can be defined as an established vocative in Turkish youth talk as it occurs 680 times (*RF*=4029.67 per million) by 30 female and 38 male speakers and the referents include both female and male addressees in the corpus. Below is a sample concordance for *kanka* in the corpus.

Left Context	Match ▽	Right Context
tamam. hadi iyi geceler	kanka	. teşekkür ederim.
eve geçince bi denersin.	kanka	şimdi artık biraz oturttun ya.
anladım	kanka	. haklısın. ya ben • ben seninle aynı fikirdeyim v
((1.0)) işte	kanka	bilmiyom. ben • aşırı böyle. buhran mı diyim? ne
olabilir	kanka	.
anladım	kanka	.
selam	kanka	. napıyosun? iyi misin?
oo!	kanka	• biraz değişikmiş bu. ((short laugh))
umarım ya. neyse	kanka	• hadi çalışmaya başlayalım. ben seni birazdan gö
benim sıram baya artmıştı. ((chuckles)) ben	kanka	yani bütün sorulara baktım. iki saatim geçmişti.
ha˙ uyanırım ya. şey yapıyorum	kanka	• yatağa gidersem uyanamam da. çalışma odamdaki k
((laughs)) ama	kanka	• en kolay çözüm o. yoksa bak • uyuyorum tamam mı
tu. sadece bi tanesinin dolabı yanmış. yan taraf/	kanka	yanındaki dolap da yanmış. nası oldu anlamadım. b
	kanka	kameranı aç. bunları atmiyim.
dal yaprak mı diyim sana	kanka	?

Figure 14 Sample concordance lines for *kanka* in the CoTY

In the analysis, kanka and its variants kank, kanki, kanks were analysed as a single cluster. Each concordance line was analysed in its expanded context individually to identify role of the vocative in discourse. The analysis revealed two broader pragmatic functions of *kanka* 'dude': (i) *organizational* functions and (ii) *interpersonal* functions.

The practices of turn management, topic management and summons are related to the ***organizational*** uses of a vocative. These uses include addressee identification/selection, interruption (i.e., turn management); launching, expanding, shifting, changing, closing, summarizing the topic (i.e., topic management); getting the interlocutor's attention (i.e., summons) in discourse.

The functions which sustain ***interpersonal*** functions include uses of humour/irony (i.e., badinage); attenuating potential threats to positive/negative face (i.e., mitigators); and

personal comments, ritual exchanges, agreements, face boosters (i.e., relational) which mainly reinforce solidarity and in-groupness (McCarthy & O'Keeffe, 2003). Table 27 shows the overall results where the tabulation of functions of *kanka* by its positions are presented.

Table 27 Functions of *kanka*

Function		*Utterance Position*				*Total*
		Stand-alone	*Initial*	*Medial*	*Final*	
Organizational	Turn Management	4	64	6	17	**91**
	Topic Management	3	135	32	49	**219**
	Summons	10	31	2	21	**64**
	Sub-total	*17*	*230*	*40*	*87*	*374*
Interpersonal	Badinage	1	50	8	24	**83**
	Mitigator	1	57	17	26	**101**
	Relational	1	61	14	46	**122**
	Sub-total	*3*	*168*	*39*	*96*	*306*
Total		***20***	***398***	***79***	***183***	***680***

Overall, pragmatic functions of *kanka* are more frequent in organizational level than that of the interpersonal level (*n*=374 and *n*=306, respectively). At this point, it is also important to note that vocatives themselves do not realize these functions on their own but rather have a supportive or signalling role in discourse (McCarthy & O'Keeffe, 2003; Wilson and Zeitlyn, 1995). Each function will be presented and exemplified below.

Among all functions, *kanka* stands out most frequently (*n*=219) in the *topic management* function in the CoTY. An example to this function is presented in extract (30) below which is from a talk on dating. Speaker is explaining how he would act if he was in his friend's shoes. While holding his turn, he refers to his interlocutor as *kanka* which marks the beginning of a series of tips he proposes.

(30) Y-2-FM-04122020

SM10004 yani. bilmiyorum. ben • ((name_male))'ın yerinde olsam. daha çok şey yaparım hani. hoşlanıyo bile olsam şey derim • kanka taktik veririm. çiçek mi gördüm • bak çiçek çok güzelmiş hani git ver. fısıldarım.
well. I don't know. if I were in ((name_male))'s shoes. I would do more things, you know. even if I like (her) I would say… kanka, I would give him tips. If I see flowers I would whisper 'look, the flowers are very beautiful, go give them (to her)'.

Turn-granting is the second-most frequently used function (*n*=91) used for organizing the talk. In (31) below, *kanka* occupies turn-initial position in the utterance and it displays turn-granting function in interaction.

(31) Y-2-F-05062021

1	SF11013	kanka bişey sorcam. annen havuza gönderir mi?
		kanka I'll ask you something. would your mom send you to the pool?
2	SF11012	((1.0)) gönderir herhalde.
		((1.0)) I guess she would.
3	SF11013	tamam. güzel. nice.
		(English)
		okay. fine. nice.

Within organizational uses, the final function is *summons* which occurs 64 times in the corpus. It is the function which most frequently occurs in turn-initial position (*n*=31) which involves either direct calls to the interlocutor to come or utterances to attract attention (Jefferson, 1973; Leech, 1999). An example to this function is (32) in which the speakers are browsing the online menu to order food and the speaker directs his interlocutor's attention to options to eat. Lexical item *bak* 'look' shows a preference to co-occur with *kanka* in summons function (*freq.* =9, *MI3*= 6.73, L3-R3). *Bak* in its 2nd person singular imperative form was previously identified to carry the functions of gaining attention and emphasizing in spoken Turkish discourse (Aksan & Demirhan, 2017; Ruhi, 2011). Vocative *kanka*, in these cases then, indicates a pre-sequence to attracting attention.

(32) Y-3-M-02122020-a

1	SM13001	kanka bak. Burger diyo. döner diyor.
		kanka look. there is Burger (King). there is döner.
2	SM11012	açık mı diyo?
		is it open?
3	SM13001	kebap. Türk mutfağı diyor. kokoreç. kumpir. kumru.
		kebab. there is Turkish cuisine: kokoreç. kumpir. kumru.
4	SM11011	Burger'dan söyleyelim. Burger'dan yiyelim.
		let's order from Burger. let's eat Burger.

Humourous interaction is a prominent feature in the CoTY and speakers make use of vocative *kanka* particularly in utterance-initial position as the set-up for the playful

utterances. Example (33) below is an instance of badinage function manifested in the interaction among two female speakers. SF10004 has an upcoming birthday and she is wondering whether the boy she likes will get her a present or not. She pesters her friend SF12013 about her anxiety over this issue iteratively in the talk. In the end, SF12013 opts for a joking response and different than the use of *kanka* in previously presented turn management function, *kanka* here is used as the opening line for the joke directed at SF10004 in the following turn.

(33) Y-2-F-04052021

1	SF10004	abi doğum günüm geldi. acaba ((name_male)) kutlicak mı doğum günümü? ***abi, it is my birthday. I wonder whether ((name_male)) will congratulate me?***
2	SF12013	şimdi kanka • bak bi bana. ***well kanka, have a look at me.***
3	SF10004	hı˙ ***yeah.***
4	SF12013	müneccime benziyo muyum? ***do I look like a psychic?***
5	SF10004	((snorts)) ya beni ciddiye al. kutlicak mı kutlamicak mı? ***((snorts)) take me seriously. will he congratulate me or not?***

Another example for the function of badinage is (34) below where three male speakers are talking about cooking. SM10002 proposes that he can come over and cook for them to which SM10017 responds with a sarcastic response.

(34) Y-3-M-07102021

SM10002	bak şimdi. siz ((name_place))'de kalırken size gelip yapabilirim gelip. ***look now. I can come over and cook for you when you are staying at ((name_place)).***
SM10017	yok kanka zehirlenmek istemiyorum. ***no kanka, I dont want to get food poisoning.***

Relational function is the most frequently observed function within interpersonal uses of *kanka* (*n*=122). In example (35) below, speakers are talking about the pandemic. SM11008 is distressed as one of his neighbours has coronavirus and he is worried that he might have been exposed to the virus as well. His friend SM11009 reassures him in (35) that there is no need to worry. In the extract, SM11009 uses vocatives *kanka* three

times in two turns. First two instances in utterance-initial positions have relational functions in which SM11009 consoles his friend and states his own opinions about the concern. The third *kanka* occurring with the expression *neyse* 'anyways' has an organizational function, it is used to change the topic.

(35) Y-2-M-05122020

1	SM11009	kanka bi şey yok lan merak etme. ***kanka, it's nothing, do not worry.***
2	SM11008	oğlum ben gelip gittim o kadar. ***man, I went there many times.***
3	SM11009	her gün mü geldin? ((laughs)) ***did you went there everyday? ((laughs))***
4	SM11008	evet. asansöre bindim. ***yes. I used the elevator.***
5	SM11009	kanka bi şey olmaz. rahat ol ya! neyse kankam. ***kanka, it'll be fine. relax! anyways, kankam.***
6	SM11008	evet? ***yes?***
7	SM11009	ıı˙ BluTV ücretsiz olmuştu ya? ***err BluTv was free, right?***

Mitigator function (*n*=101) is observed in contexts of potential or directed threats at face. For instance in (36) below, a group of three male speakers are talking about romantic relationships. Among the speakers, SM11005 has a long-term relationship. Other speakers, SM11006 and SM11007, are shocked to hear that SM11005 has been dating his girlfriend for over six years. SM11006 poses a potentially face threatening act but mitigates his imposition on negative face of SM11005 by utilizing vocative *kanka* in the utterance-initial position.

(36) Y-3-M-07122020-a

1	SM11006	oha˙ oğlum! kaç yaşında başladın çıkmaya başladın lan? ***whoa man! how old were you when you started dating?***
2	SM11005	((2.0)) on iki. ***((2.0)) twelve.***
3	SM11006	yuh anasını satıyim! ***whoa what the!***
4	SM11007	((2.0)) on bir! ***((2.0)) eleven!***
5	SM11006	kanka sıkılmıyon mu peki böyle aynı insandan? ***kanka I mean don't you get bored of the same person?***
6	SM11005	yok aga. ***no bro.***

The present analysis regarding the pragmatic functions of *kanka* shows that the range of functions a single vocative type can exhibit in interaction. Young speakers of Turkish use *kanka* for both organizational and interpersonal purposes, the most frequently utilized functions are topic management and relational functions in the CoTY.

4.3.2.3.6 Address shifts: *bro* versus *kardeş*

The overview of distribution of tokens presented above indicate that the vocatives are extensively used in the corpus, in other words, 88% of the speakers (*n*=108) in the CoTY use at least one type of nominal vocative in their interactions. A total of 15 vocatives (*çocuğum, dostum, evladım, ezikler, göt, güzelim, hatun, hayatım, hocam, ibne, kuzu, millet, moruk, reis, şekerim, tatlım*) exist as idiosyncratic uses due to the fact that there is only one user for each of these vocatives in the whole corpus. Nevertheless, there are also frequent instances of multiple vocatives used by a speaker to address the same interlocutor. In the analysis, these instances are marked as *address shifts* which affirm the dynamic, responsive and goal-oriented nature of interaction among friends. In the CoTY, majority of speakers (72%) use more than one type of vocative (*M*=4.4, *SD*=3.04) in a single conversation and the maximum number of vocative types a speaker uses is 15 (*n*=2). In order to explore this particular phenomenon, vocatives within the semantic category of sibling and their patterns of uses were examined in their expanded context.

As was previously mentioned in 4.3.2.3.4, it was identified that a number of young speakers of Turkish who use vocative *bro* (*n*=5) use the vocatives *abi*, *aga, kardeş* for the same interlocutors. Considering that these vocatives are semantically related (i.e., *abi* means elder brother, *aga* is a contemporary slang form of elder brother, and *kardeş* means sibling in Turkish) do they exhibit the same pragmatic functions in discourse? Do the shifts in addresses occur randomly or is there an underlying interactional goal for using *bro* over *abi, aga* or *kardeş* in particular contexts in Turkish youth talk? To explore these questions, pragmatic functions of *bro* and *kardeş* were identified and compared. Tokens *aga* and *abi* were excluded from the analysis as both are derivatives of *ağabey* 'elder brother' which marks its referent with the feature of seniority of age as compared to *brother* in English which does not denote any age-based seniority between the addresser and the addressee.

In the CoTY, *bro* occurs 38 times while *kardeş* occurs 58 times (see Table 28). Though the number of users of *kardeş* is limited (*n*=26, 20 male and 6 female speakers) compared to the users of *bro* (*n*=5, 3 male and 2 female speakers), the results illustrate a tendency of difference in the distribution of pragmatic functions they have.

Table 28 Comparison of functions of *bro* versus *kardeş* as vocatives

Voc.	*No of organizational functions*			*No. of Interpersonal functions*			*Total*
	Turn Management	*Topic Management*	*Summons*	*Badinage*	*Mitigator*	*Relational*	
bro	4	13	5	3	5	8	**38**
kardeş	3	8	5	20	15	7	**58**

As shown, the functions of *bro* focus on organizational functions (58%) while *kardeş* is utilized mainly for interpersonal functions in in discourse (72%). These vocatives are gender-inclusive, they are used by and for both males and females in the CoTY.

Functional difference is better illustrated when focused on the address shifts of same speakers in a single conversation. In the CoTY, there are two male speakers who use both *bro* and *kardeş* for the same addressee in their talk. In both instances, speakers use *bro* for organizational functions (turn and topic management) and they switch to *kardeş* for the function of badinage. To present the phenomenon in more detail, the shift from *bro* to *kardeş* by one of these speakers in a single conversation will be presented.

SM11010 is a 17-year-old male speaker from Çanakkale and engages in talk with his classmates in extracts (37-a) and (37-b). The speakers firstly talk about the national university entrance exam and SM11010's friend asks for help to find the government website which has official information regarding the faculties and universities in Turkey. SM11010 provides him with the answer, his friend thanks him, and in the following turn, SM11010 uses *bro* to close the topic:

(37-a) Y-3-2M1F-14052021

SM11004 buldum. okay. YÖKAtlas. dil. thank you!
found it. okay. YÖKAtlas. language. thank you!

SM11010 valla bro. öyle yani.
well bro. that's it.

Later in the course of their chat, the speakers SM11004 and SM11010 start talking about social media and SM11010 shares his negative opinions regarding social media platform TikTok and its users. In (37-b), SM11010's friend says that even though he does not use TikTok often, he sometimes uses it to only edit photos and post them in his private account. To build on this, SM11010 switches to vocative *kardeş* in his pre-sequence to an episode of banter in turn 2. SM11010 teases his friend for using TikTok in a light-hearted manner accompanied by laughter. His interlocutor does not get the joke, thus in the following turn 4, SM11010 explains the underlying context of the joke and reconstructs it in a discourse of camaraderie. To this, his friend responds with laughter as a sign of equilibrium of understanding for the implicated meaning and affirming the in-groupness in turn 7. The successive laughters at the end of the episode also nurtures the solidarity and intimacy (Coates, 2007; Everts, 2003) among interactants.

(37-b) Y-3-2M1F-14052021

1	SM11004	onları yapması eğlenceli oluyo işte. size de attımdı ya. prive hesabıma da attımdı. işte okulda fotoğraflarınızı koydum. ***it is fun to do those (things). I sent them to you, too. I also posted them on my private account. I put your photos at school.***
2	SM11010	kardeşim. benim gözümde tamam mı? o saçma hareketleri yapmadığın sürece • varsın. ((short laugh)) ama yaparsan • da • arkadaşlığımızı değerlendirebilirim. ((short laugh)) ***my brother. to me, you know? as long as you do not do those stupid things, you are in. ((short laugh)) but if you do, then, I might consider our friendship.***
3	SM11004	anlamadım. ***I did not get it.***
4	SM11010	ya hani böyle atıyolar ya kendilerini böyle. bişi yapıyor böyle. artist artist giriyolar böyle. birbirini falan dövüyolar. o hareketleri <yapmadığın sürece • >/1> ***you know they throw themselves like this. they do stuff like this. they do dramatic entrances. they beat eat other or something. <unless you do those things >/1>***
5	SM11004	<haˑ >/1> ***<ah >/1>***
6	SM11010	dewamke yani. destekliyorum. ***so, dewamke I mean. I support you.***
7	SM11004	((laughs))
8	SM11010	((laughs))

The shift from *bro* to *kardeş* suggest that even though these vocatives appear as semantically equivalent lexical items in English and in Turkish, the speakers show

preference for using them for different pragmatic functions in interaction. Echoing Zwicky's observations (1974), Leech (1999) underlines that vocatives mark speaker-referent relationship and the vocatives are never sociopragmatically neutral. The shifts of address, in this sense, show that these sociopragmatic characteristics are not static but rather responsive to the interactional goals a speaker aims to achieve.

Though earlier studies either focused on addressee-identifying and summoning role of vocatives (Biber et al., 1999; Leech, 1999; Schegloff, 1968) or the selection of vocatives determined by the power relationship between the speakers (Brown & Gilman, 1960; Brown & Ford, 1961; Oyetade, 1995); the recent studies adopt a more sociopragmatic approach and highlight that vocatives operate on "the interpersonal space" (Jworski & Galasiński, 2000, p. 79) and thus they can function as pragmatic markers which encode self-positioning (Heyd, 2014), face concerns (Rendle-Short, 2007; Tsakona & Sifianou, 2019) and politeness (Afful, 2006; Clancy, 2015; Formentelli, 2007; Wood & Kroger, 1991). The overview of nominal vocatives in the CoTY with regard to addressers, addressees, forms, positions, semantic categories, pragmatic extension and address shifts indicated that young speakers of Turkish attend to both organizational and interpersonal needs they experience in interaction with their friends. The results corroborate the perspective that the repertoire of vocatives is extensively used to project and enhance the intimate level of relationship the speakers have by attending to face concerns and creating playful and humorous tone in interaction.

In brief, this section presented the types, the distributions, the patterns of vocatives followed by a focus on the formal characteristics and pragmatic functions of the register-specific vocative *kanka*, as well as the presenting an account of the phenomenon of address shifts in the data. Moving on from vocatives, the next section of this chapter presents the findings with regard to *vague expressions* used among Turkish speaking young people in the CoTY.

4.3.3 Vague expressions

As a defining characteristic of interaction between close familiars, vagueness expressions are reported to have the power "to project co-constructed worlds" (Clancy & McCarthy, 2015, p. 444) through creating a space of meaning-making among the speakers. Vague language is made up of words and expressions which refer to non-

specific or generic items in an imprecise way (Channell, 1994). There is an a priori assumption that vague language is a characteristic of talk between close associates as they exhibit a high involvement speech style and the closer the relationship the more vagueness is manifested (Clancy, 2016; Evison et al., 2007; Stenström et al., 2002). As a result, vague language has been extensively explored as a typical feature of youth talk. The following section will introduce the labels and categorizations that have been used for the linguistic particles of vague language along with the development of approaches utilized to treat these particles.

4.3.3.1 Defining vague expressions

Vague language covers linguistic particles in various forms and these particles go by various terms in the literature such as 'set-marking tags' (Dines, 1980), 'generalized list completers' (Jefferson, 1990), 'extension particles' (DuBois, 1993), 'vague category identifiers' (Channell, 1994), 'general extenders' (Overstreet, 1999; Overstreet & Yule, 1997), 'coordination tags' (Biber et al., 1999), 'discourse extenders' (Norrby & Winter, 2002), 'particles with vague reference' (Aijmer, 2002), 'placeholders' (Halliday, 2003), and 'vague category markers' (O'Keeffe, 2004) and This study adopts the umbrella term 'vague expressions'.

Channell (1994) provides the first comprehensive study on vague language and categorizes vague language into three broad groups for British English. These include inherently vague words or phrases (e.g., *things*); vague additives such as vague approximators (e.g., *around*) and tags (e.g., *and stuff like that*); and vague quantifiers for amounts, numbers, frequency and likelihood (e.g., *loads of, sometimes*). Adding on Channell's (1994) categories, Cutting (2007) proposed additional types of vagueness expressions which are vague lexis (i.e., metonymies), vague reference (i.e., anaphoric nouns and adverbs, indefinite pronouns), and vague clausal or utterance-level features.

Several other studies have revised Channell's (1994) framework as well, among them there are also cases which adopts a more minimal approach to group the vague expressions into two: namely vague language (e.g., *things, like, kind of, or something, I think*), and approximations (e.g., *around, or so, about*). Overstreet's (1999) seminal study focuses on a specific linguistic particle within vague language, namely general extenders which she defines as expressions which are non-specific (thus 'general') and extend

grammatically complete utterances (thus 'extenders'). She further provides a non-exhaustive list for potential general extenders in American English which consists of adjunctive general extenders (expressions beginning with *and*) and disjunctive general extenders (expressions beginning with *or*). Overstreet's work is important to illustrate that vague language exhibit multifunctionality which is manifested more in interpersonal functions compared to referential functions.

Earlier works on vague language explored the referential meanings of the forms used while the most recent work reveals the relational management maintained by these linguistic particles. Cutting (2007) documents that initial studies on vague language in the 1960s to 1980s focused on implicitness (e.g., Garfinkel, 1967; Grice, 1975; Gumperz, 1982) and identified it as a feature of informal conversation (e.g., Lakoff, 1972). In the 1990s, Carter and McCarthy (1997) highlighted the interpersonal function of mitigation achieved by vague expressions and Channell (1994) proposed that vague language marks the 'shared knowledge' of speakers and Cutting (2001) underlined that vague language affirms in-group membership and fosters solidarity while at the same time othering the out-group. Cutting's studies (2001, 2007) are important in the sense that context is noted as a significant variable for the use of vague expressions. In this line, Overstreet (2012) differentiates two approaches to analyse vague expressions. The first approach involves the analysis of vague expressions as part of sentence meaning using semantic analysis frameworks while the most recent second approach examines vague expressions as part of utterance meaning through pragmatic analysis often utilizing corpus methods.

The recent treatments of vague language which investigate the use of vague expressions in the relational domain also adopt a more cross-linguistic perspective. As an example, Overstreet's (2005) comparative analysis of general extenders among American English and German speaking adults indicated that even though the forms used were different in the formal level they are similarly used to mark the assumptions of being similar, informative, accurate and polite. For both languages, the study lists the functions of intersubjectivity, solidarity, iconicity (through reduplicated forms), evaluation (when used with pejorative nouns) for adjunctive extenders while the functions of (lack of) accuracy, negative politeness, emphasis are identified for disjunctive general extenders.

There are also studies which investigate the variation in the use of vague language within a single language. Vaughan et al. (2017), for instance, deals with vague category markers as turn-final items in Irish English from LCIE and in British in English from CANCODE. Adopting a corpus pragmatics approach, a predetermined list of vague expressions is analysed to test whether they trigger speaker change in interaction. The results showed that vague category markers occur more frequently before speaker change in British English.

Similarly for general extenders, Aijmer (2013) uses ICE data and reveals variations in the forms of general extenders used by American, Australian, British, and New Zealand English speakers (e.g., *and stuff* is used more frequently in American, Australian and New Zealand English while British English has *and things* as the counterpart). The study also notes functional differences between adjunctive and disjunctive general extenders and indicates that *and*-extenders facilitate in-group membership and social similarity/establish familiarity, similarity and solidarity by avoiding explicitness while *or*-extenders are used for hedging as they express tentativeness or assertation that the content could be inaccurate.

4.3.3.2 Brief overview of related work on vague expressions

The overview of studies on vagueness expressions in corpus-based spoken discourse will be briefly presented in two sub-sections: existing work on Turkish and foci of youth language research.

4.3.1.3.1 Vague expressions in Turkish

Studies on Turkish vague language do not exhibit a systematic and coherent investigation of the issue, rather there are studies which either briefly mention vagueness as a linguistic phenomenon while handling other linguistic analyses, along with a couple of studies which explore specific vagueness expressions individually.

Among the studies which analysed vagueness expressions under the overlapping terminology of discourse/pragmatic/interactional markers, *şey* 'thing' is among the most widely investigated linguistic particle. As a preliminary and comprehensive investigation of the issue, Özbek's (1995, 1998) works which explored Turkish discourse markers

based on naturally occurring spoken data from speakers between the ages 23-50 define the primary function of *şey* as a discourse marker used for planning/organizing the speech by the speaker. Building on the initial comparative findings concerning Turkish *şey* and English *well* in his master's thesis, Yılmaz (1994, 2004) highlights the multifunctionality of *şey* in spoken Turkish and identifies its functions in both conversational structure domain and interpersonal domain. Yılmaz's (2004) study is prominent in the sense that it also defines *şey* as a placeholder used for vagueness in interaction. Erdoğan's (2013) study utilized the STC data and identified the functions of *şey* as self-repair, introducing a new topic, holding the floor and signaling a topic shift, as well as working as a face-saving device in spoken Turkish interaction. In Furman and Özyürek's (2007) study in which the researchers take a more developmental perspective, they compared the speech of 3-, 5-, and 9-year-old children with that of adults to explore the interactional aspects of Turkish spoken discourse. The results indicated change in the frequency and functions of discourse markers and *şey,* which is identified to have a function of 'nominal filler' in interaction and as the marker which is acquired the earliest.

There are also a number of studies which report discursive and pragmatic observations regarding lexical item *falan* which corresponds to a range of English general extenders such as *or so, and all, or whatever.* It is reported as a multifunctional lexical device in contemporary informal spoken Turkish (Özgen & Karataş, 2018; Tekin, 2015). Yet the functional properties of this vague expression remain unexplored.

4.3.1.3.2 Vague expressions in youth talk

One of the earliest observations regarding vague language in youth talk is found in the study of Labov (1982) who states that vagueness can be associated with power relations and presenting oneself 'in' a group among English speaking adolescents. Indeed, the majority of work exploring vagueness focused on the interpersonal functions of vague expressions. Corpus-based studies investigating the pragmatic aspects of vague expressions reported that these linguistic items construct comradeship and solidarity among young speakers of English and Spanish (Stenström et al., 2002; Stenström, 2005; 2014). Adopting a comparative approach to youth talk, Palacios Martínez and Núñez Pertejo (2015) investigated placeholders in English youth talk in the COLT and LIC compared to data of adult speakers in the DCPSE and the BNC. The results indicated that

youth talk showed a larger repertoire of placeholders. The study argues that placeholders have both interpersonal functions such as insult, comradeship, attitudinal functions and discourse organizational functions such as holding the floor. Comparing the COLT data with the DCPSE, Palacios Martínez (2011a) indicated that general extenders occur more frequently in adult talk yet some specific forms are found to be more frequent in youth talk. In a complementary study, Palacios Martínez (2011b) identifies the most frequently used placeholder in the COLT as *thing(s)* and its variants and while the most commonly used general extenders are identified as *and stuff, and everything*, and *and that* which are used for expressing self-connection and reaffirmation of group membership.

There is also a second batch of studies which questioned the factors influencing the use of vague language in youth talk. Among them, one of the most frequently investigated parameter is social class. Stubbe and Holmes's (1995) variationist study on pragmatic markers in Wellington Corpus of Spoken New Zealand English (WCSNZE) data mentions that 'set marking tags' *sort of/kind of* are used more frequently in young middle-class females and young working-class males which are defined as groups 'associated with leading language change from below'. Research on the COLT also reported that some forms of vague language occurred more frequently in the speech of speakers from specific social classes (Stenström et al., 2002).

Additionally, Cheshire (2007) showed variety of use for the forms of general extenders by speakers from different social classes in her study based on interview data of English speaking 96 adolescents between ages 14-15. The following studies, though, showed that rather than gender and social class, the use of specific vague expressions depends on the context (Adolphs & Carter, 2013; Andersen, 2001; Cheng & O'Keeffe, 2015; Clancy, 2016; Koester, 2007). In this line, Tagliamonte and Dennis's study (2010) tested the social, grammatical and discourse-pragmatic factors in relation to the use of general extenders in spoken Canadian English and revealed that even though general extenders were more frequent in youth data, socioeconomic status was not a differentiating factor for the use of vague language.

4.3.3.3 Findings: Vague expressions in the CoTY

In this section, findings regarding vague language in the CoTY data will be presented. This study focused on two categories of vague expressions: vague references and vague additives. In order to identify the related lexical particles in the corpus, a list of candidate vague expressions was generated. While forming this list, existing literature as well as major reference works on Turkish grammar (Banguoğlu, 2011; Göksel & Kerslake, 2005; Lewis, 2000) were consulted. Additionally, the emergent list of related tokens compiled during corpus construction stage was integrated. For each candidate vague expression in the list, separate KWIC analyses were conducted.

Taking a pragmatic approach to the study of vague expressions as was suggested by Overstreet (2012), the first 10 concordance lines and their expanded contexts for each candidate vague expressions were qualitatively investigated to determine whether the expression exhibits vagueness as a part of utterance meaning. The identified list of vague expressions was categorized using a revised version of Channell's (1994) categorization.

In the following sub-sections, the distribution of vague references and vague additives in the CoTY will be presented, the patterns and functions in the data will be exemplified and discussed. Following that, the relationship between communicative purposes within a conversation and the use of vagueness expressions will be investigated. For this purpose, the most frequently occurring vague additive *f(a)lan* will be examined based on the conversational communicative purposes of discourse units in the data.

4.3.3.3.1 Types and distribution

The analysis yielded 26 types of 4438 tokens of vague expressions in the corpus. These expressions were grouped under two main categories: vague references and vague additives.

Table 29 below lists the types of tokens, total number of tokens retrieved from the corpus (*TN*), the absolute frequencies of total number of identified vague expressions (*AF*) along with the relative frequencies (*RF*) per million in a descending order.

Table 29 Distribution of vague expressions

Vague expression category	*Sub-category*	*Token*	*Gloss*	*TF*	*AF*	*RF*
vague references	**indefinite reference**	şey	*thing*	3538	**2093**	12403.11
		biri(si/leri)	*somebody, one (of them)*	355	**273**	1617.80
		başka(sı/ları)	*another (one)*	166	**166**	983.72
		hepsi	*all*	109	**109**	3827.80
		kimse	*none*	95	**95**	562.97
		aynısı	*the same one*	30	**30**	177.78
		diğer(ler/i)	*the other one*	28	**28**	165.93
		birbiri	*each other*	26	**26**	154.08
		(bir)çoğu	*most (of them)*	43	**21**	124.45
		bazı(sı/ları)	*some (of them)*	54	**20**	118.52
		kimi(si/leri)	*some (of them)*	17	**10**	100.74
		hiçbiri(si)	*none*	13	**13**	77.04
		öteki(si)	*the other one*	9	**5**	53.33
		herhangi biri	*any (of them)*	3	**3**	17,78
		tümü	*all*	2	**2**	11,85
	generic reference	insan	*one*	315	**21**	124.45
		adam	*man*	382	**5**	29.63
	Sub-total				*2920*	*10865.20*
vague additives	**approximators**	neredeyse	*almost*	22	**22**	130.37
		civarı	*around*	3	**3**	17.78
		-(i)ms(i)	*-ish*	4	**2**	11.85
		gibi	*around*	549	**1**	5.93
		tahminen	*around*	1	**1**	5.93
	general extenders	f(a)lan	*and stuff, or anything*	1468	**1468**	5462.37
		m-	*and stuff, or something like that*	5171	**16**	94.82
		vesaire/vs	*and others, or anything*	3	**3**	17.78
		f(a)lan	*and stuff, or anything*	2	**2**	11.85
	Sub-total				*1518*	*5648.41*
Total					***4438***	*16513.61*

TN: Total number of tokens in corpus, *AF:* Absolute frequency, *RF*: Relative frequency per million

As presented in Table 29 above, vague references occur more frequently in the corpus (*AF*= 2920, *RF*= 10865.20) than vague additives (*AF*=1518, *RF*=5648.41). The most frequently occurring vague expression is placeholder *şey* 'thing' in the whole corpus (*AF*=2093, *RF*=12403.11) followed by general extender *f(a)lan* 'and stuff' overall. In following sections, each category of vague expressions will be presented in more detail.

4.3.3.3.2 Vague references

Identified indefinite and generic references under the category of vague references are presented in Table 30 below.

Table 30 Vague reference tokens tabulated by frequencies and speakers

Sub-category	*Token*	*Gloss*	*No. of tokens*		*No. of speakers*					
			AF	*RF*	*All*		*Female*		*Male*	
					N	*%*	*N*	*%*	*N*	*%*
indefinite reference	şey	*thing*	**2093**	12403.11	113	92	60	97	53	87
	biri(si/leri)	*somebody, one (of them)*	**273**	1617.80	83	67	47	76	36	59
	başka(sı)	*another (one)*	**166**	983.72	70	57	45	73	25	41
	hepsi	*all*	**109**	3827.80	49	40	33	53	16	26
	kimse	*none*	**95**	562.97	49	40	27	44	22	36
	aynısı	*the same one*	**30**	177.78	18	15	12	19	6	10
	diğer(ler/i)	*the other (one)*	**28**	165.93	25	20	19	31	6	10
	birbiri	*each other*	**26**	154.08	21	17	13	21	8	13
	(bir)çoğu	*most (of them)*	**21**	124.45	49	40	28	45	21	34
	bazı(sı/ları)	*some (of them)*	**20**	118.52	16	13	8	13	8	13
	kimi(si/leri)	*some (of them)*	**10**	100.74	8	7	4	6	4	7
	hiçbiri(si)	*none*	**13**	77.04	12	10	9	15	3	5
	öteki(si)	*the other one*	**5**	53.33	4	3	1	2	3	5
	herhangi biri	*any (of them)*	**3**	17.78	2	2	2	3	0	0
	tümü	*all*	**2**	11.85	2	2	0	0	2	3
generic reference	insan	*one*	**21**	124.45	19	15	15	24	4	7
	adam	*man*	**5**	29.63	4	3	0	0	4	7
Total			***2920***	*10865.20*	*123*	*100*	*62*	*100*	*61*	*100*

AF: Absolute frequency, *RF*: Relative frequency per million

A few words on vague references in Turkish is due here. In Turkish, vagueness is expressed in many linguistic levels, this study focuses on referential status of lexical items within its scope of analysis. Noun phrases have four referential status which are definite, indefinite, categorical, and generic reference in Turkish (Göksel & Kerslake, 2005). Among these, indefinite noun phrases can denote specific entities or non-specific entities. In this line, this study treats non-specific indefinites and generic references as the sub-types of vague references in Turkish.

The results show that the most frequently occurring indefinite reference is ***şey* 'thing'** which is often defined as a placeholder item in both research on Turkish and English. It can replace a word as well as a whole clause in Turkish (Göksel & Kerslake, 2005). The analysis shows that referent of *şey* 'thing' could be present in the same utterance, within the local context, the extended context, or the referent may not be present at all. In all cases, the interaction continues as the speakers share a common understanding of the issue.

Below in excerpt (38), 18-year-old high school graduates from İzmir talk about a movie they both watched. SM13002 criticizes the movie's storyline and comments on what one of the characters did in the movie. The referent of *şey* is revealed when speaker repairs himself in turn 1.

(38) Y-2-M-21112020

1	SM13002	onun arkadaşlarının öldüğünü şey yaparak ıı˙ kamerada görmesine rağmen • gidip şeyin içine girmesi. ***even after doing the thing that her friends are dead umm seeing in the camera, her entering into the thing.***
2	SF13003	evet ya. ***yeah.***
3	SM13002	akıl hastanesinin içine girmesi. hani daha aptalca bi karakter motive'i olabilir mi? ***entering into the asylum. I mean, could there be a dumber character motive?***

It is observed that within the same turn, SM13002 uses *şey* for the second time. As SF13003 already knows what he refers to, she answers *evet ya* 'yeah'. In the following turn 3, SM13002 reveals what second *şey* in turn 1 refers to. In this case, the referents of *şey* are present in the local context. It is important to note that even though use of vagueness expressions is considered to have the potential to violate the cooperative

principle (Grice, 1975; Overstreet, 1999; Overstreet & Yule, 1997), the results show otherwise. SM13002 carries the conversation even though he does not adhere to the maxim of quality[36], the examination of expanded context shows that this lack of precision is due to the apparent mutual investment in the experience by the speakers. As a result, the use of vagueness expressions does not disrupt the interaction, but rather facilitate it.

Sometimes the local context does not reveal the referent. An example is illustrated in (39) below where two 14-year-old friends from Kırklareli are talking online. In this excerpt, SF09004 instructs her friend to accomplish a task on her behalf. The referent of *şey* 'thing' is not present within the turn, neither in the following turns. When the extended context is examined though, it is revealed that *şey* refers to a document she mentioned earlier in the conversation.

(39) Y-2-F-02122020

SF09004 önce birinci şifreyi dene. sonra ikinci şifreyi dene. ve sana attığım şeyi atarsın. bu arada ıı jpg dosyası olarak yazıyordu. jpg dosyası böyleymiş. yani normal fotoğraf dosyası.
firstly, try the first password. then try the second one. and send the thing I sent you. by the way, umm it was written as a jpg file. that is what a jpg file is. I mean a regular photo file.

There are also cases when the referent of *şey* 'thing' is never present even in the extended context of the whole episode of conversation. In such cases, it is clear that *şey* refers to an entity or a concept within a shared conceptual space among speakers. Excerpt (40) below is an example to such a case where thing as in 'send a song thing' refers to a social media trend in which a user shares an Instagram story in their account and asks their followers to recommend them a song to listen to. Speaker SM10005 comprehends what *şey* 'thing' refers to as he comments on the procedure she had to follow and offers another advice to his friend, as well.

[36] The Cooperative Principle assumes that speakers and listeners act cooperatively to achieve effective communication in a conversation (Grice, 1975). Among its four maxims, **maxim of quality** requires speaker to provide adequately truthful information.

(40) Y-3-2M1F-16052021-c

1	SF10015	ya bi ara şey yaptım. işte şarkı yollayın şeyini paylaştım story'de. böyle duygusal şarkılar falan atmış. ama yazmıyo. ***well I did this thing. I mean I shared the 'send a song thing' in the story. she sent emotional songs and stuff. but no writing.***
2	SM10005	yakın/ yakın arkadaşlara yapcaksın. ee· DM evet hayır tarzında bişey yapıcaksın. ***you have got to do this for close friends. umm you should do something like 'DM yes no'.***
3	SM10006	DM evet hayır mı? mute'e alıyorum. acı gerçekler! sus ya! ***DM yes no? I'm muting you. the bitter truth! just shut up!***

Generic reference ***insan*** corresponds to the generic pronoun 'one' in English. In the CoTY, speakers use generic reference *insan* to refer to individuals without naming them but the referent is always denoted in somewhere else in the context. The results show that the use of generic reference is always present in declarative sentences which conveys evaluation regarding a behaviour. As a result, it is usually used in emotionally-laden contexts.

In excerpt (41) from a conversation between two 16-year-old classmates from Ankara, the main topic of the talk is that SF10011's mother meddling in the way she dresses. Overall in the talk, SF10011 argues that her mother should not intervene with her life as she does not behave like a model and morally intact grown-up.

(41) Y-2-F-18052021

1	SF10011	mesela geçen gün dedikodu yapıyolar annanemle. Ramazan ayı bi de. işte bana o kadar dini şeyden bahsediyo annem • bunu yapma. bunu yapma. anneme dedim o an. konuşurlarken. anne niye dedikodu yapıyonuz? milleti çekiştiriyonuz? dedim. günah değil mi dedim. sonra annem bana böyle böyle baktı • sana ne? dedi. beni azarladı falan. sonra ordan annanem dedi ki • hani kız haklı dedi. biz niye dedikodu yapıyoz? dedi. kapatak gitsin falan dedi. ***for example, the other day they were gossiping with my grandmother. and it is the month of Ramadan. my mother was telling me all those religious things: 'do not do this. do not do this'. I said to my mom at that time. while they were talking. 'mom why are you gossiping? talking behind people?' I said. 'isn't not a sin?' then she looked at tme like this. 'it is not your business' she said. she scolded me and stuff. then my grandma said***

'well, the girl is right. why are we gossiping?' she said. 'let's cut it out' she said.

2 SF10012 vay be.
wow.

3 SF10011 işte! cık· <u>insanın</u> başkasını yargılamadan önce kendisine bakması gerek. ((2.0)) ve kimsenin artık giyinişini hiçbi şekilde sormulamıyorum ben.
see! <u>one</u> needs to check themselves before judging others. ((2.0)) and I do not question how anyone dresses anymore.

As an argument, she shares with her friend that her mother and grandmother are often gossiping about other people. In turn 1, after reenacting an episode of such an event, she says *<u>insanın</u> başkasını yargılamadan önce kendisine bakması gerek* '<u>one</u> needs to check themselves before judging others'. Though this statement does not have a definite subject and object, the local context indicates that generic reference *insan* 'one' refers to speaker's mother and indefinite *başkası* 'other' refers to herself. SF10011 openly criticizes the behaviour of her mother in the re-enactment yet in the conclusion statement, she refrains from explicitly referring to her mother or even using the indefinite pronoun *o* 'she'. In this case, she intensifies her criticism by adding the emphasis that all proper people need to behave like that. By use of generic reference, the speaker declares and underlines a personal opinion which is an evaluative statement oriented towards the behaviours of an absent other.

The referent of *insan* 'one' can also be present in the immediate context, among the participants of interaction as exemplified in (42) below. This particular conversation takes place in Istanbul among three 16-year-old school friends while they are commuting to their school via subway. SM10007 shares with his girlfriend and his friend that he will be participating in a sailing cup in Bosphorus. His girlfriend SF10018 asks questions about the event but SM10007 is not able give any details about the cup.

(42) Y-3-2M1F-31102019

1 SF10018 nerden başlıyo? nerde bitiyo?
where does it start? where does it end?

2 SM10007 güzel bi soru. bilmiyorum.
nice question. I don't know.

3 SF10018 bizim sahilden geçiyo musunuz? geçmezsiniz herhalde.
are you passing by our coast? you don't I guess.

4 SM10007 sanmıyorum.
I don't think so.

5 SF10018 caddeden belki geçer.

		maybe it passes by the main road.
6	SM10007	geçmez ((name_place))'den geçiyoruz herhalde. oraya o kadar şey yapmışlar. ***it does not. I think we are passing by ((name_place)). they did all those stuff there.***
7	SM10008	cık˙ ordan geçer. ***no it passes by there.***
8	SF10018	tamam bakarım. ay! uf! ben konuşamadım. bilmiyorum ya! <u>insan</u> beraber kayıt yaptırır! ***okay I'll check that. ah! I couldn't talk. I don't know! <u>one</u> registers together!***
9	SM10007	ya ben • isteyerek mi <u>şey</u> sanıyosun? ***well I. do you think I do that on purpose?***
10	SM10008	((XXX))
11	SF10018	ama kayıt yaptığında diyceksin ki • ((name_ SF10018)) ben yaptım. böyle bişe var. ***but when you register you should say ((name_ SF10018)) I did it. there is such a thing.***
12	SM10007	haa˙ seni gördüğüm mü var? Allah Allah! beş gün oldu görüşmeyeli! ***do I even see you? for God's sake! we haven't seen each other for five days!***
13	SF10018	bi görüşmedik diye! ***just because we haven't seen each other just once!***

In this excerpt, in turn 8, SF10018 snaps at her boyfriend for not telling her about the event earlier. She says *<u>insan</u> beraber kayıt yaptırır*! '<u>one</u> registers together!' in which the generic referent *insan* is used to emphasize the expected code of behaviour which her boyfriend did not conform to. In the following turn 9, it is evident that SM10007 is well aware that *insan* 'one' refers to himself, thus he tries to confront the criticism by explaining that he did not do it on purpose. Notice that he also integrates vagueness expression *şey* 'thing' in the same turn, *ya ben isteyerek mi <u>şey</u> sanıyosun?* 'do you think I do that (lit. <u>thing</u>) on purpose?' which refers to the act of improper behaviour his girlfriend previously implied by the utterance *<u>insan</u> beraber kayıt yaptırır*! '<u>one</u> registers together!'. The use of generic reference, then, does not create any ambiguity in interaction as the referent of *insan* acknowledges that the criticism, thus the face threat, is directed at him and responds her with an offensive counter strategy.

4.3.3.3.3 Vague additives

Vague additives are lexical items which accompany or are attached to noun phrases to convey imprecision in meaning. As Table 31 shows, this category includes approximators and general extenders as the sub-categories for the identified vague additives (*AF*=1518, *RF*=5648.41) in the corpus.

Table 31 Vague additive tokens tabulated by frequencies and speakers

Sub-category	*Token*	*Gloss*	*No. of tokens*		*No. of speakers*					
			AF	*RF*	*All*		*Female*		*Male*	
					N	*%*	*N*	*%*	*N*	*%*
approximator	neredeyse	*almost*	**22**	130.37	18	15	10	16	8	13
	civarı	*around*	**3**	17.77	3	2	3	5	0	0
	-(i)ms(i)	*-ish*	**2**	11.85	2	2	0	0	2	3
	gibi	*around*	**1**	5.92	1	1	1	2	0	0
	tahminen	*around*	**1**	5.92	1	1	0	0	1	2
general extender	f(a)lan	*and stuff, or anything*	**1468**	8699.4	99	80	54	87	45	74
	m-	*and stuff, or smt like that*	**16**	94.81	12	10	2	3	10	16
	vesaire, vs	*and others, or anything*	**3**	17.77	3	2	2	3	1	2
	mıdır nedir	*or whatever*	**2**	11.85	2	2	0	0	2	3
Total			***1518***	*5648.4*	*123*	*100*	*62*	*100*	*61*	*100*

AF: Absolute frequency, *RF*: Relative frequency per million

The results show that the most frequently occurring approximator is ***neredeyse* 'almost'** (*AF*=22, *RF*=130.37) in the corpus which is often used to give a rough estimate regarding quantities or states. In (43) below, for instance, the topic of conversation is the economy. 16-year-old SF09007 and 17-year-old SF1101 from Sakarya are talking about how the commodities they used to afford became much more expensive. SF09007 compares the current price of her study table with its former price from two years ago. In turn 4, her interlocutor SF11011 expresses her astonishment by noting the discrepancy between the prices via approximator *neredeyse* 'almost'.

(43) Y-2-F-14052021

1 SF09007 altta fiyatı • iki yüz kırk sekiz lira yazıyo. ya da üç yüz kırk sekiz. şu anki fiyatı yedi yüz lira.
the price below says two hundred and eight liras. or three hundred forty eight. İts current price is seven hundred liras.

2 SF11011 şaka gibi.
like a joke.

3 SF09007 iki senede değişmiş.
it changed in two years.

4 SF11011 iki katına çıkmış neredeyse.

it is <u>almost</u> twice as much.

5 SF09007 aynen iki katından da fazla.
exactly, more than that.

Vague additives in Turkish data are linguistically manifested in bare forms as in *neredeyse* in excerpt (43), or they can be affixes as the approximator ***–(i)ms(i)*** which corresponds to *-ish* or *like* in English (*AF*=2, *RF*=11.85). This particular derivational suffix is attached to nominal roots and forms adjectives which express a degree of similarity to the entity the root nominal denotes in Turkish. Excerpt (44) below shows the lexical manifestation of this particular approximator within the lexical item *ekşimsi* 'sour-ish' by a speaker to guess the taste of his interlocutor's recipe for taco. In this conversation, SM10002 makes an offer to his friends that he can prepare tacos for them and starts explaining the dish to his friends. In turn 1, SM10011 intervenes and becomes the co-teller of the narrative of describing the dish. His description though, is hypothetical, yet he manages to enter a similar conceptual space with SM10002. In turn 4, he further guesses the taste as *tatlı* 'sweet' but SM10002 corrects the guess as *acı* 'bitter'. This contradiction is a potential threat that can leave SM10011 out of the shared conceptual space, so in turn 6, SM10011 makes use of the approximator *–(i)ms(i)* in his utterance to hedge his previous statement and converges with SM10002. This way, it is clear that he manages to stay at the common conceptual space with his friend as SM10002 replies with the reduplicated response token *evet evet evet evet evet evet* 'yes yes yes yes yes yes' in an overlapped turn in the following turn.

(44) Y-3-07102020

1 SM10002 ((laughs)) ve abi bak. onun tadını düşünemiyorum. sana yapılışını izletirim. onun tadının güzelliğini düşünemiyorum!
((laughs))) and look bro. I can't even imagine the taste of it. I'll show you how it is made. I can't image how delicious it must be!

2 SM10011 ağzına atıyosun. et yumuşacık lokum gibi. bi de böyle lif lif.
you put it in your mouth. the meat is as tender as a delight. and the texture is like fibrous.

3 SM10002 evet evet evet. lif lif.
yes yes yes. fibrous.

4 SM10011 ağzına atıyosun. tatlı.
you put it in your mouth. sweet.

5 SM10002 tatlı değil. acı.
not sweet. bitter.

6	SM10011	ekşimsi acı tatlı. bissürü <tat birlikte oluyo. >1> ***sourish bitter sweet. lots of <flavors together. >1***
7	SM10002	<evet evet evet evet>1> evet evet! ***<yes yes yes yes >1> yes yes!***
8	SM10013	o baharatlar! ***those spices!***
9	SM10002	ve şey böyle. o taze soğanın şeyi var böyle. sululuğu ve kıtırlığı böyle. kırt! diye böyle. ağzında hissediyosun. ***and like. there is that fresh onion thingy. the juiciness and the crunchiness, you know. just like that. you feel it in your mouth.***
10	SM10011	ah! biz daha öğrenciyiz. yapma böyle! ***ah! we are just students. don't be like that!***

Research on vague additives in English have so far mainly focused on general extenders (Aijmer, 2013, 2015; Cheshire, 2007; Tagliamonte & Denis, 2010; Levey, 2012; O'Keefe, 2004; Overstreet, 1999, 2005; Pichler & Levey, 2011). In addition to studies on English, studies which explored French (DuBois, 1993), German (Overstreet, 2005), Spanish (Palacios Martínez, 2011a, 2011b, 2011c) and Swedish (Winter & Norrby, 2000) differentiate adjunctive-disjunctive distinction for general extenders. Adjunctive general extenders are vague expressions which have noun phrases followed by expressions beginning with *and* such as *and stuff*, *and everything, and all that*; and disjunctive general extenders which are followed by expressions beginning with *or* such as *or whatever, or something like that, or what.* The results indicate that, this particular distinction is not extensively observed in Turkish youth talk. Moreover, identified general extenders are used as adjunctive and disjunctive general extenders interchangeably with the exception of ***vesaire* 'et cetera'**. Based on formal characteristics, only the expression *vesaire* 'et cetera' could be defined to correspond to the adjunctive form 'and others' in English.

In excerpt (45) below, for instance, *vesaire* 'et cetera' is used by 16-year-old female speaker from Kastamonu. This particular vagueness expression is the closest equivalent to general adjunctive extender *and stuff* in English both in terms of its formal structure and pragmatic function.

(45) Y-2-F-03122020

1	SF11002	eşit ağırlık da seçebilirdim. sayısal da seçebilirdim. ee˙ dil de seçebilirdim. eşit ağırlığı seçebilirdim. çünkü geneller genelde denemede eşit ağırlığım daha ağır basıyodu yani • sıralamam daha önde oluyodu. ee˙ babam • şey diyodu hani istersen sayısalı yaparsın diyodu. ona bakılırsa işte hani dil de biliyodum. ve ilgim vardı. bilmiyorum. hep • hep arasındaydım. ***I could have also chosen maths-literature. I could have also chosen science. I could have also chosen the language. generally, my rankings were higher for maths-literature. umm my father said 'well, you can do science if you want'. but similarly I was good at the language, too. and I was interested. I don't know. I was always, always in between.***
2	SF11001	hı-hı˙ ***mm-hmm.***
3	SF11002	hani bi şeyim çok iyi değildi hani ondan biraz bundan biraz. o yüzden hani ne seçeceğime karar veremiyodum. ondan sonra ((name_female)) hocanın ısrarlarıyla • ve ee˙ o çocuğun bizim okula gelip tekrardan şey yapmasıyla vesaire. ***it was not like I was very good at something. I was good enough in each of them. that's why I was not able to decide which branch to choose. then with the insistence of Teacher ((name_female)) and umm that guy coming to our school again and doing stuff et cetera.***
4	SF11001	((chuckles))
5	SF11002	ıı˙ o akşam işte babamla konuştum. dedim ki • hani böyle böyle geldi. hani artık ne seçeceğimi bilmiyorum kafam daha da çok karıştı falan. sonra öyle olunca işte uzun bi konuşma yaptıktan sonra dil seçmeye karar verdim. ***err that evening, I talked to my father. I said 'you know, he (that guy) came to our school. I don't know what to choose anymore, I'm even more confused.' then, after a long conversation, I decided to choose the language.***

In the conversation above, SF11002 is talking about the process of how she determined her track (language) in high school. She mentions that she had been confused and unable to make a decision in turn 1, and SF11001 responds with the non-lexical response token *hı-hı* 'mm-hmm' to encourage her interlocutor to continue. In turn 3, SF11002 indicates that she came to a decision after talking with a teacher and meeting a person (referred as *o çocuk* 'that guy' in turn 3) who was studying literature at college. SF11002 had talked about the visit of this person previously in the earlier parts of the conversation, thus in this section of the talk, the speaker avoids giving details about it as she talked about this issue with her interlocutor before. This way, the referential underspecification achieved through the use of vague expression *vesaire* operates within the principle of cognitive

economy as Schwarz-Friesel & Consten (2011) proposes. The vagueness expression replaces the utterance or a series of utterances and as a result, redundancy is avoided. It is interesting to note that speakers may even shorten vagueness expressions as in (46) below where short form ***vs* 'etc'** is used which is the abbreviation used for *vesaire* in written Turkish. The speaker enunciates each of the letters separately as /ve/ for letter 'v' and /se/ for letter 's'. A total of two speakers use vs in their speech in the CoTY, and among them one of them use it as a truncated form of the vague expression *vesaire*.

(46) Y-3-M-06122020-2

1	SM12013	bi de şey Monopoly gibi bişey oynuyordu. ha˙ o bak Monopoly zaten şey • ka/ kutu oyunlarında falan oynanıyo. yani toplanılınca. ***and he was playing something like Monopoly. oh look Monopoly is already played as a board game. I mean when people get together.***
2	SM12012	hıı˙ ***hmm.***
3	SM12013	zaman geçiriliyo bi şekilde. ***the time is being spent somehow.***
4	SM12012	evet. öyle dediğim gibi. Cyber Park'a bayağı yükseldim ben. bekliyom. bakalım. çıkış haftası da tam vize haftama geldi. ***yes. just like I said. I am hyped up for Cyber Park. I'm waiting. we'll see. its release date overlaps with my midterms.***
5	SM12013	aa! ***oh!***
6	SM12012	çok güzel olduğunu söyleyemem ama • ayın onuna kadar • sınav vs her şey bitmiş oluyo zaten. ((XXX)) ***I can't say that it is very good but till the tenth of this month, the midterms et cetera will be over anyway.***

There are only two instances of general extender *mıdır nedir* 'or whatever' in the corpus and both are used in utterances with negative semantic prosody. Research on disjunctive general extenders (Aijmer, 1985; Overstreet, 2005) identify the functions of lack of accuracy, negative politeness, emphasis yet in excerpt (47) below, 17-year-old SM11004 uses general extender *mıdır nedir* in the interrogative utterance *((name_male)) mıdır nedir çocuğun adı?* 'the name of the boy is ((name_male)) or whatever?'. In this case, general extender *mıdır nedir* does not function as a token to get affirmation regarding an uncertainty but as a token to check whether the interlocutor has the background knowledge regarding the person SM11004 plans to talk about.

(47) Y-3-2M1F-14052021

1	SM11004	şey vardı ya. bu zeki diyodu/ diyolardı. tarihten falan yüz almıştı. o zamanlar çok bayağı şey yapmışlardı. konuştular. ((name_male)) mıdır nedir çocuğun adı? ***remember the thing. they said that he was smart. he got a hundred points in history or something. they did a lot of things back then. they talked. the name of the kid is ((name_male)) or whatever?***
2	SM11010	haˑ evet. ***ah yes.***
3	SM11004	haˑ o • mesela en fazla yine şeymiş işte. otuz vermiş hoca. ***oh him. they say the maximum is said to be the thing. teacher gave thirty.***
4	SF11008	yok artık! ***no way!***
5	SM11010	kanka o çocuk var ya. sana bişe söyliyim mi? ((name_male)) kanka • hiç çalışmıyo. çalışmayan bi çocuk kanki. ***kanka, that kid, you know. let me tell you something. kanka, ((name_male)) does not study at all. he is a kid who never studies.***

In the following turns in (47) above, it is evident that SM11004 knows the name of the person as he shares more information about him in turn 5. In this case, then, general extender *mıdır nedir* serves a relational purpose of marking an attitude about a person or an event and conveying this attitude to the interlocutor in interaction.

So far, approximators *neredeyse* 'almost', *–(i)ms(i)* '-ish' and general extenders *vesaire/vs* 'et cetera', and *mıdır nedir* 'or whatever' were presented and exemplified as vague additive tokens in the CoTY. Now, the focus will be on another form of general extender observed in Turkish through a process of affixation. Identified as a process of generating general extenders, ***m*-reduplication** is a form of reduplication process observed in Turkish. It is formed by repeating a word with replacing its first consonant with *-m* (e.g., *kalem malem* 'pencil or something like that') or attaching *-m* if the word starts with a vowel (e.g., *iyi miyi* 'good or something like that'). Göksel and Kerslake (2005, p. 91) notes the function of m-reduplication as "generalizing the concept denoted by a particular word or phrase to include other similar objects, events, or states of affairs" and Gencan (2007) adds that the *m*-reduplicated component represents indefiniteness. Echoing these observations, the analysis shows that this particular structure functions to form general extenders in the corpus (*AF*=16, *RF*=94.81). As exemplified in (48-a & b & c)

below, speakers use *m*-reduplicated general extenders which correspond to *or-plus* (48-a, 48-b) and *and-plus* (48-c) extenders.

(48-a) Y-2-F-05122020-1 [topic: the experience of an earthquake]

SF12008 biraz sonra dedim • ben ders çalışıcam ama ders mers çalışmadım.
later, I said I'll study but I didn't study or anything like that.

(48-b) Y-3-M-06122020-1 [topic: a football match]

SM12011 kavga mavga çıktı ya o zaman.
remember there was a fight or something like that then.

(48-c) Y-3-M-02122020-b [topic: American tv series Punisher]

SM12012 aynen‿.aynen. yok ediyo ortalığı. tarıyo marıyo. giriyor. tek başına mekan basıyo.
exactly exactly. he terminates everything. he opens fire and stuff. he enters. he invades the place.

Corpus data also shows that speakers apply *m*-reduplication to English words as observed in excerpt (49) from a conversation about online games among three 18-year-old friends from Mersin. In this case, SM12012 forms *m*-reduplication by repeating English word *update* with adding the consonant *-m* to it. As a result, the cluster *update mapdeyt* 'update and stuff' is generated in which lexical item *mapdeyt* is the ortographic representation for *m*-reduplicated form of *update*.

(49) Y-3-M-06122020-2

1 SM12014 ya benimki kaldırır mı bilmiyorum. bakalım.
well I don't know whether my computer will run the game. we'll see.

2 SM12012 indir log'la dene bi. indirirse oyna.
download it and try it with the log. play if it downloads.

3 SM12013 ha˙
ah!

4 SM12012 ama ((XXX)) yetmiş GB'mış.
but ((XXX)) it is seventy gigabytes.

5 SM12014 onu indirmek de sıkıntı.
downloading that is a hassle as well.

6 SM12012 ama update mapdeyt dahil değil. update gelirse kaç GB gelir…
but update and stuff are not included. how many gigabytes would be an update…

Overall, the analysis shows that the identified forms of general extenders in Turkish function to extend grammatically complete utterances. Despite that, the results also demonstrate that it is not possible to categorize Turkish general extenders into two distinct groups of adjunctive or disjunctive expressions as it is in English. Turkish general extenders can semantically correspond to either of the both groups of general extenders. Furthermore, their forms are not restricted to *and-plus* and *or-plus* formulas generated for general extenders in English.

To conclude this section on vague expressions in the CoTY, the final part of the analysis will present a more detailed account of the most frequently occurring general extender and the second-most frequently observed vague expression ***f(a)lan*** **'and stuff'** (*AF*=1468, *RF*=8699.36) in the whole corpus. This particular form of general extender will be presented in more detail in the following section with a particular focus on its use with regard to its communicative purposes within interaction among Turkish speaking youth.

4.3.3.3.4 Communicative purposes and vague language: *f(a)lan*

This final part illustrates an approach to identify the communicative purposes of vague expressions by making use of a taxonomy of conversational discourse types (Biber et al., 2021; Egbert et al., 2021). The particular focus will be on general extender *f(a)lan* 'and stuff'. First, its formal characteristics and patterns in the corpus will be presented. Then, the distribution of communicative purposes will be presented and exemplified.

General extender *f(a)lan* is linguistically manifested as *falan (AF*=1441*), felan* (*AF*=17) and *filan* (*AF*=10) in the corpus. The initial observations indicated that general extender *f(a)lan* is more multifunctional than what is prescribed for its 'traditional' usage in Turkish. It is an extensively used vague expression in the CoTY as it is present in 92% of conversations in the corpus and the majority of speakers in the whole corpus (88%) use it across all ages. Table 32 provides details regarding the use of *f(a)lan* tabulated by speaker ages.

Table 32 Distribution of *f(a)lan* by speaker age

Age	*No. of speakers*	*% of age group in CoTY*
14	10	83
15	8	73
16	34	81
17	24	100
18	23	96
Total	*99*	*80*

In terms of its positions in the utterances, *f(a)lan* is not observed as a stand-alone utterance neither it occupies utterance-initial position in the corpus. As a general extender, it is typically found in utterance-medial and utterance-final positions. Interestingly, utterance-final positions make up of only 28% of the positions (*n*=406) while the majority of tokens occur in utterance-medial (*n*=1062, 72%) positions in the data. Below are sample concordance lines for *f(a)lan* in utterance-final position from the corpus.

Left Context	Match	Right Context
bu da kabul etmedim diye geri şey yapıyo • çekiyo	falan	. bazı böyle şeyler oluyo. ((short laugh))
ayınlıyolar. böyle kaç? bin dokuz yüz doksanlarda	falan	. belki seksenler. filmi çekiyolar. yayınlıyolar.
işte zorsa zor. imkansız olmadığı sürece yaparız	falan	. ben baktım abi böyle. ((short laugh)) dedim ki •
le adam gösteriyo. diyo • burası bizim kampüsümüz	falan	. ben böyleyim • aha! beni de alın!
altı bölüm yayınlamış adam.	falan	. ben gördüm. annemle oturup izli/ izlicez ee˙ diy
dokuz on	falan	. ben sizden yüksek yaptım.
me? hani böyle küçük küçük birsürü soru var. test	falan	. ben şu an altıncı testi bitirdim bugün. e dün üç
ızlı iletildi. aktarımda hiç bi problem yaşamadım	falan	. benim baya egom kabardı!
m oluyo! diye. ben orda koymuşum kafamı yatıyorum	falan	. berbat bişeydi ya! bi de on ikinci sınıfta direk
ne kadar? yedi bin	falan	. beş bin.
altı yayınlanmicak da • şimdi dörtte o kadar genç	falan	. beşinci sezonda çok genç sadece saçları beyazlam
i • yok sırtım ağrıyo. yok tam rahat oturamıyorum	falan	. bi de ben otururken bağcık kurmayı çok seviyorum
teşekkür ederiz. kal/ git/ işte katıldığınız için	falan	. bi de şey yaptılar. onların altına • event'i • t
inler üstünden dedim. sonra diğerlerini kaldırdım	falan	. bi o kaldı. üzerinde de üç tane karınca vardı. g
yolar ormana falan filan. kız bir şeyler uyduruyo	falan	. bi şeyler oluyo. ıı˙ zincir sesi falan çıkartıyo

Figure 15 Sample concordance lines for *f(a)lan* in the CoTY

Corpus identifies two collocates of *f(a)lan*, namely *filan* and *fişman*. These lexical items simply 'extend' the general extender *f(a)lan* further. The chunk *falan filan* occurs 58 times in the corpus and is used by 21 unique speakers (*freq.* =58, *MI3*= 18.33, L3-R3). The other chunk *falan fişman*, on the other hand, appears to be idiosyncratic within the sample as only a single speaker uses this expression (*freq.* =9, *MI3*= 13.18, L3-R3). Excerpts (50-a&b) are examples for such uses. Excerpt (50-a) below is from an episode

of storytelling in which 17-year-old female speaker from Kırıkkale is telling her interlocutor what a friend of hers said about a girl they saw in a park.

(50-a) Y-2-F-05062021

SF11012 pazartesi ben ((name_female))'le ((name_male))'yla buluştum ya parkta. hani işte biz oturuyoduk. ((name_male)) şeye dedi. benim arkam dönüktü. ((name_male)) şey dedi • işte arkamızdaki kız pişti oynuyo ne güzel falan filan dedi. öyle normal konuşmaydı.

on Monday, I met ((name_female)) and ((name_male)) in the park. well, we were sitting. ((name_male)) said. my back was turned. ((name_male)) said 'the girl behind us is playing cards, how nice and stuff'. it was such an ordinary conversation.

Above in (50-a), SF11012 does not repeat what her friend said verbatim but rather rephrases the utterance in a loose way. She conveys the message that her friend commented that he liked how the girl they saw was playing cards. By integrating *falan filan* 'and stuff', SF11012 expresses that her friend added similar comments about the girl, as well. General extender *falan filan* 'and stuff' invites SF11012 interlocutor to be on the same conceptual ground with her.

In (50-b) below exemplifies the use of the other chunk *falan fişman* 'and stuff'. In this exerpt, a 17-year-old female speaker from Eskişehir is narrating what she did with her cousins the previous day. She notes that they ate some 'things' *bişeler,* exemplifies one of those things as potato chips, and inserts *falan fişman* to indicate that they also ate some other snacks. In this case, *falan fişman* refers to other varieties of the previously mentioned snack (i.e., potato chips).

(50-b) Y-2-F-20052021

SF11009 ondan sonra şey yaptık. oturduk bi yerde. bişeler • cips • falan fişman yedik.

then we did this thing. we sat somewhere. things, we ate potato chips and stuff.

Moving from the formal characteristics of *f(a)lan* to its pragmatic functions, the preliminary observations regarding vague references and vague additives in sections 4.3.3.3.3 and 4.3.3.3.2 showed that these vagueness expressions are employed for various relational functions such as (i) showing attitude, (ii) conveying evaluation about

a person, (iii) expressing emotion, (iv) converging with interlocutor, and a number of (v) discourse-organizational functions such as avoiding redundancy and establishing the discursive flow. Studies so far also underlined that vague expressions mark an assumption of shared knowledge or co-conception between speakers and thus establish social closeness (Channell, 1994; Cheshire & Williams, 2002; Overstreet, 1999; Overstreet & Yule, 1997, 2002), convey attitudes and feelings (Overstreet, 2012), mitigate face threatening acts (Aijmer, 2013; Overstreet, 1999), carry discourse-organizational functions such as holding the turn or signalling turn exchange (Aijmer, 2013; Winter & Norrby, 2000). While the scholars indicate that the use and the frequency of vague expressions depend on situational context and topics (Overstreet, 1999; Cheshire, 2007), the contextual environment of the vagueness expressions is generally examined by taking a wider lens into the context such as the register types. Cheng (2007), for instance, examined spoken academic, business, conversation and public genres in Hong Kong Chinese (HKCSE) and native-English speaker (NES) corpus data. Not surprisingly, vague expressions were found to be used more frequently in genre of conversation which is the most informal discourse type, followed by business, public and lastly academic discourses. The results are not surprising as scholarly work suggests that vague language is often manifested in informal and intimate discourses as these domains have speakers which possess a wider shared knowledge base. 'Informal' or 'intimate' discourse; however, presents a broad domain of talk in which various communicative purposes can be integrated by the speakers.

At this point, it is important to note a couple of preliminary observations regarding the instances of vague language in the CoTY data. Even though the corpus data belongs to the single register of informal communication between friends, the situational contexts vary immensely. Furthermore, the speakers engage in multiple communicative goals within a single episode of conversation in the CoTY. As previously presented in Chapter Three, each conversation in the CoTY was also tagged for a number of speech events (e.g., conflict talk, gossip talk, troubles talk, storytelling, among others) at macro level (i.e., whole conversation was assigned tags). The macro level annotation of speech events was implemented as there was no readily available framework to identify linguistic boundaries of the speech events in spoken informal conversations. Still, this macro level annotation yielded the observation that general extender *f(a)lan* is inclined to appear more frequently in specific speech events, and in case of the pilot analysis, it was the conversations which included storytelling episodes. While this observation confirms

that it is vital to examine the local and situational context of vague expressions in order to investigate their communicative purposes, there have not been any study to systematically investigate the use of vague expressions across distinct speech events or any other defined units of discourse in a single register. For this purpose, this sub-section of the current chapter aims to provide a systematic account of investigation for revealing the patterns of functions general extender *f(a)lan* exhibits in distinct and coherent units of communication. In order to identify the distinct speech events systematically within spoken corpus data, a framework developed by Egbert et. al (2021) was implemented. Egbert et al.'s (2021) method provides a corpus-based procedure to identify functionally coherent and sequentially bounded sequences of utterances which are operationally defined as Discourse Units (DU) and describe the communicative purposes of these units. The framework was chosen on the basis that it was developed using a sample of informal and face-to-face conversational interactions from a spoken corpus (i.e., the BNC Spoken 2014). The observation that a single register of informal communication among friends can be segmented into smaller episodes with distinct and potentially multiple communicative purposes was also reflected in the underlying assumptions for developing the aforementioned framework. The framework is pioneering in the sense that it proves wide implications for systematically analysing discursive and pragmatic patterns within distinct functional units in a single register as well as across-registers.

For this purpose, the analysis followed Egbert et al.'s (2021) methodology to identify and categorize conversational discourse units and then assign dominant communicative purposes to these discourse units. The definition of a discourse unit indicates that it (i) has an identifiable beginning and end, (ii) is coherent in terms of a major communicative goal, and (iii) has a minimum of five utterances of 100 words[37]. Within this definition then, an identified single discourse unit has the potential to include multiple occurrences of a token of interest, which is the token *f(a)lan* in this study. In this line, each instance of *f(a)lan* (*n*=1468) was examined in its expanded context and the boundaries of discourse units which contained the vague expression *f(a)lan* were marked. The identification of boundaries revealed that a single discourse unit can include multiple tokens of *f(a)lan* as previously assumed. The results yielded 1206 discourse units in total. Later, these discourse units were manually coded for nine communicative purposes using Egbert et al.'s (2021) framework.

[37] It should be noted that the taxonomy is developed based on spoken English data.

Communicative purposes of discourse units in this framework are: (1) situation-dependent commentary, (2) joking around, (3) engaging in conflict, (4) figuring-things-out, (5) sharing feelings and evaluation, (6) giving advice and instructions, (7) describing or explaining the past, (8) describing or explaining the future, and (9) describing or explaining (time neutral). Taking into account Biber et al.'s (2021) argument that a single discourse unit can have multiple communicative purposes but only has one dominant purpose, the final coding for each discourse unit highlighted only its dominant purpose[38]. The results provided the types of communicative purposes accomplished in identified conversational discourse units in which general extender *f(a)lan* is used at least once by at least one of the interlocutors. Table 33 below shows the overview of communicative purposes of discourse units in which *f(a)lan* occurs in the corpus.

Table 33 Overview of communicative purposes of DUs in which *f(a)lan* is used

Communicative Purpose	*No. of DUs with f(a)lan*	*No. of conversations*
[FEL] sharing feelings and evaluations	**376**	44
[PAS] describing/explaining the past	**359**	42
[DES] describing/explaining (time neutral)	**196**	40
[FTO] figuring-things-out	**135**	32
[JOK] joking around	**43**	24
[SDC] situation-dependent commentary	**38**	22
[FUT] describing/explaining the future	**35**	24
[ADV] giving advice and instructions	**15**	8
[CON] engaging in conflict	**9**	5
Total	***1206***	*48*

Based on the existing literature, it was hypothesized that vague language was saliently used in conversations referring to shared experiences. It was also indicated that vague expressions were used for conveying opinions, attitudes, emotions. The above presented analysis confirms this observation and shows that the most salient communicative goals of discourse units in which general extender *f(a)lan* occurs in the corpus are *sharing feelings and evaluations* (*n*=376), followed by *talking about past events* (*n*=359). In the following part, different types of communicative purposes of the discourse units in which

[38] Please see Biber et al. (2021) and Egbert et al. (2021) for the detailed methodology for identifying discourse units and dominant communicative purposes in corpus data.

f(a)lan occurs will be presented in a descending order of frequency as presented in Table 33 above. For each communicative purpose type, excertps from the CoTY will be presented and explained. Owing to space constraints, the focus will be on the most salient functions of *f(a)lan* intertwined with these communiative purposes.

Sharing personal feelings and evaluations

Discourse units which exhibit the communicative purpose *personal feelings and evaluations [FEL]* include expressing emotions and personal opinions as well as conveying personal evaluations. In the following excerpt (51), a discourse unit with the dominant communicative purpose of conveying emotions is presented. In this excerpt, speakers are two 16-year-old female friends from Adana and the prevailing emotion is the feeling of longing.

(51) Y-2-F03122020-2

1	SF11004	okulu • özledim sanırım! ***I think I missed the school!***
2	SF11003	ben de okulu özledim. ***I missed it, too.***
3	SF11004	yani • bunu pek diyeceğim aklıma gelmezdi ama. ***well, I never thought I'd say this, but.***
4	SF11003	((XXX)) hiç özlemedim. ***I didn't miss ((XXX)).***
5	SF11004	hı-hı˙ dersleri değil zaten • ortamı ortamı özlüyoruz. ***mm-hmm not the lessons. we miss the atmosphere.***
6	SF11003	aynen okul arkadaşlarımı falan çok özledim. ***exactly. I missed my friends and stuff a lot.***
7	SF11004	aynen. ***exactly.***

In turn 6 in excerpt (51) above, SF11003 indicates that she misses her school friends and other things related to school. As her interlocutor SF11004 shares the same feeling, she replies with the engagement token *aynen* 'exactly' in turn 7. The response token is used to give the message that she understands how SF11003 feels and that she misses the school and her friends as well. This way, general extender *falan* connects SF11003 with SF11004 in the shared affective domain.

While Biber et al. (2021) and Egbert et al. (2021) include feelings, evaluations, opinions, personal perspectives, and beliefs within the scope of the communicative purpose of *sharing feelings and evaluations [FEL],* the analysis on *f(a)lan* revealed that young speakers of Turkish make use of a high number of *f(a)lan* particularly within episodes of gossip talk (*n*=148, observed in 40% of FEL discourse units) in the corpus. As a form of evaluative talk, gossip talk is oriented towards an absent other (Eder & Enke, 1991). In the CoTY, the results show that 'the other' can be a person whom speakers personally know or a well-known public figure. In all cases, gossip talk enables speakers to negotiate a stance towards the other (Jaworski & Coupland, 2005). In discourse units with the communicative purpose of *sharing feelings and evaluations,* the analysis of the episodes of gossip indicate that young speakers of Turkish use *f(a)lan* for facework as gossiping has the potential to be a face threatening speech act (Blum-Kulka, 2000; Guendouzi, 2001; Thornborrow & Morris, 2004). To exemplify, the discourse unit of *sharing feelings and evaluations* in excerpt (52) below presents an episode of gossip talk in which 17-year-old speakers from Kırıkkale are engaging in evaluative talk about a mutual friend.

(52) Y-2-F-05062021

1	SF11012	bişey diyim mi? ((name_female))'yla yakın gibiler biraz. ***let me tell you something. she is a bit close with ((name_female)).***
2	SF11013	hmm˙ ***well.***
3	SF11012	bahsetmiştim ya sana da. yani biraz özelini falan da biliyo gibi. ***I told you this earlier, too. I think she knows about her private life and stuff.***
4	SF11013	neyi biliyo gibi? ***knows what?***
5	SF11012	özelini. ***her private life.***
6	SF11013	hmm˙ bilmiyom. ya ona bişey anlatılmaz ona bişey anlatılırsa bütün Kırıkkale'ye yayılma ihtimali var. ***well I don't know. one shouldn't tell her anything. if you share something with her, she'll spread the news to the whole province.***

In excerpt (52) above, SF11012 initiates the gossip talk in turn 1. By initiating the gossip talk, she threatens her positive face, thus, SF11012 makes use of *falan* as a hedging device both to handle this threat and to invite SF11013 to display an affiliative stance with her. To respond to this call, SF11013 poses a question in turn 4 in order to show alignment with SF11012 and encourage her to expand her narrative. By making use of general

extender *falan,* then, the speakers are able to avoid threats to positive face and establish a shared stance which fosters in-groupness among the speakers.

Describing or Explaining the Past

The analysis revealed that 30% of discourse units in which *f(a)lan* is used are about *describing or explaining the past [PAS]* (*n*=359). Within these discourse units, speakers either reminisce about the events they experienced together or one of the speakers narrate a personal experience which her/his interlocutors hears for the first time. For the first case, the analysis shows that the fact that the narration is based on a shared experience makes the explicit descriptions redundant, thus speaker integrates general extender *f(a)lan* into the narration as exemplified in (53) below.

(53) Y-2-F-03122020-1

1	SF11002	evet◡ .aynen. ◡bi de orda Türklerle karşılaşmıştık hatırlıyo musun? **yeah exactly. and we met Turkish people there, remember?**
2	SF11001	hı-hı˙ ***mm-hmm.***
3	SF11002	biz orda şey hani bağırıyoruz • ((name_female1)) ((name_female2)) şuraya buraya diye ((laughs)). **we are like screaming '((name_female1)) ((name_female2)) this way that way' ((laughs))**
4	SF11001	((chuckles))
5	SF11002	sonra • aa˙ siz de mi Türksünüz • falan oldular. ***then they were like 'ah, are you Turkish, as well?' and stuff.***
6	SF11001	bi de tavrımızdan falan anlamış. ***and they said that they got it from our attitude and stuff.***
7	SF11002	((short laugh))

In excerpt (53) above, two 16-year-old speakers from Kastamonu are recalling their memories of a school trip abroad. Both speakers SF11001 and SF11002 makes use of falan in the turns 5 and 6 when they talk about a shared experience in the past, respectively. The use of *falan* in turn 5 should be particularly noted as it contributes to the construction of an episode of reenactment[39] (after Sidnell, 2006) in interaction. The results show that younger speakers of Turkish frequently makes use of *falan* in

[39] **Reenactment** is the representation or depiction of a previously occurring event, often drammatically, in interaction.

reenactment within the discourse units with the communicative purpose of *describing or explaining the past* in the CoTY unit (*n*=50, observed in 14% of PAS discourse units). It is also noteworthy that the speakers make use of *falan* as a formule of *f(a)lan ol-* 'to be f(a)lan' (*freq*. =8, *MI3*= 10.117, L3-R3) in the CoTY. In this chunk, the verb *ol-* 'to be' can be inflected for tense/aspect/modality and person as in the expression *siz de mi Türksünüz falan oldular* 'They were like 'ah, are you Turkish, as well?' and stuff' in turn 5. The analysis indicates that young speakers of Turkish use this formule as a discursive strategy to construct reenactments in talk.

In addition to the conversations about the past in which the speakers recall an event they experienced together as previously exemplified in (53), speakers also use *f(a)lan* when they narrate an intimate or a personal experience which the listener hears for the first time. An example to this is presented in (54) below where 16-year-old speaker SM10004 from İzmir is talking about his private life with his close female friend SF11006.

(54) Y-2-FM-04122020

1	SM10004	yani • şeydim böyle. ben orda • demiştim içimden. tatlı kızmış. falan. <demiştim. >1> ***well, I was like, I told myself 'she is cute' and stuff. <that's what I said.>1>***
2	SF11006	<hıı˙ >1> ***<oh. >1>***
3	SM10004	ama hani • hoşlanma yok. sadece tatlı kız. sonra • kamp olayları falan başlayınca • hafiften • şey oldu böyle. hmm˙ daha tatlı kız. ***but no liking. just a cute girl. then when the camping event started, it slightly became a bit like 'well a very cute girl'.***
4	SF11006	((chuckles)) daha tatlı! ***a very cute girl!***
5	SM10004	aynen. daha tatlı. yani öyleydi. ama • hani • o • kamp olaylarında zaten şey • parti olaylarında zaten gözüm • o sıralarda onda değildi. ((XXX)) ***exactly. very cute. I mean that was how it was. but you know, during those camping events, party events I was not actually interested in her back then. ((XXX))***
6	SF11006	((short laugh)) aga/ aga be! aga be! ***come on bro! come on bro!***

In the conversation above, SM10004 reveals what he thought when he saw 'the girl' in the past through a group of segmented utterances *yani şeydim böyle* 'well, I was like'; *ben orda demiştim* içimden 'there I told myself'; *tatlı kızmış* 'she is cute'; *falan* 'and stuff' which make up of turn a single turn of 1. This segmented narration suggests that

SM11006 is sharing an emotion-laden and private story with his friend. Following the utterance *tatlı kızmış* 'she is cute' in which the SM10004 shares his experience of starting to develop feelings towards that person, he inserts general extender *f(a)lan* at the end of his utterance for face concerns. As confessing an intimate story has the potential to threaten SM10004's self image, he mitigates his message via *f(a)lan.* SF11006's use of non-lexical response token *hıı* 'ah' in turn 2 encourages SM10004 to continue and affirms that he handled maintaining his positive face.

Describing or Explaining (Time Neutral)

Communicative purpose of describing or explaining things in time-neutral space is another cluster (*n*=196) of discourse units in which *f(a)lan* is used. This particular communicative purpose includes the episodes of talks on facts, information, people and events without specifing the time of occurrence. As was previously indicated in section 4.2 where topical and lexical characteristics of the corpus data were presented, the physical appearance and their daily routines are among the conversation topics among young speakers of Turkish. As a result, one of the most frequently observed communicative purpose for the discourse units with *f(a)lan* is *describing or explaining (time-neutral*) *[DES]* as exemplified in (55) below.

(55) Y-2-F-18052021

1	SF10011	ben saçımı kıvırcık yapsam kıvırcık oluyo. çok rahat şekilleniyo. ***if I fix my hair curly it stays curly. it gets styled easily.***
2	SF10012	senin saçın düz gibi. daha çok. ***your have straight hair. relatively.***
3	SF10011	düz gibi. dalgalı gibi. ama ben şey gördüm bi vidyodan. ***it is like straight. like wavy. but I saw something in a video.***
4	SF10012	ama daha çok düz. ***but it is more like straight.***
5	SF10011	hep böyle düz dalgalı zannedenler kıvırcık çıkıyomuş böyle. onların şampuanlarından kremlerinden sürünce • onlar asıl şeklini alıyomuş. ***those who think their hair is straight or wavy but they actually come out curly. when you use shampoo or conditioner for curly hair, your hair takes its original shape.***

6	SF10012	benim saçım genelde şu tarafı düz oluyo. buraya gelince • bi böyle böyle bişeyler oluyo falan. ***my hair is usually straight on this side. when it comes to this part, it becomes something like this or something like that and stuff.***
7	SF10011	((laughs))

Within the discourse unit in excerpt (55) above, 17-year-old female speakers in Ankara are talking about their daily hair care routines. While SF10012 is explaining the hair styling problems she has, she uses *falan* in her utterance *benim saçım genelde şu tarafı düz oluyo. buraya gelince • bi böyle böyle bişeyler oluyo falan* 'my hair is usually straight on this side. when it comes to this part, it becomes something like this or something like that and stuff' in turn 6. In this case, vague expression *f(a)lan* is used to depict the usual physical characteristics of an entity, SF10012's hair, in a humourous manner. In the following turn of 7, SF10011 responds with a laughter as she aligns with SF10012's playful depiction of her hair.

Figuring Things Out

Another communicative purpose of the discourse units in which *f(a)lan* is present is *figuring things out [FTO]*. The analysis shows that most of the time, the speakers in the CoTY are trying to figure out issues within the domain of education (*n*=102, observed in 75% of FTO discourse units). The topics the speakers try to figure out within these discourse units generally belong to future oriented issues such as school work and exams. Excerpt (56) is an example for a discourse unit with this communicative purpose. In the excerpt below, 17-year-old speakers from Çanakkale are trying to figure out procedures related to school.

(56) Y-2-FM-14052021

1	SM11004	şeyleri napıcaz? ***what do we do about the things?***
2	SF11008	neyleri? ***which things?***
3	SM11004	dersleri. sözlüler falan verilmicekmiş galiba. e· ː verilmicekse bizim mesela/ pardon. bizim derslere ze de gerek yok. ***the courses. they say there won't be any oral exams. e are no oral exams, we do not need to attend the .***
4	SF11008	gerek yok da işte belli olmaz yine onlara. ***there is no need, but you never know.***

5	SM11004	bence pazartesi günü soralım. hocalara. ***I think we should ask the teachers on Monday.***
6	SF11008	aynen͡ .aynen. ***exactly exactly.***
7	SM11004	nedir ne değildir diye. ona göre girelim. aynen. <boş :mek değil de… >1> ***to understand what it is about. we can attend the based on that. <attending all for naught… >1>***
8	SF11008	<boş yere gitmeyelim. >1> galiba bi dilekçe falan ***<no need to attend the class for naught. >1> I guess*** ***it a letter or something.***

Within the abovepresented discourse unit, SM11004 expresses in turns 1 and 4 that he does not want to attend the classes if class participation will not earn them any marks for their final grade. He is not sure about the school procedure and regulations related to this issue, thus he proposes a strategy to figure this issue out. He tells his friend SF11008 that they ask their teachers on Monday. His friend SF11008 uses the reduplicated lexical response token *aynen aynen* 'exactly exactly' to show her aggreement with this solution. In addition to accepting SM11004's suggestion, SF11008 also shares the piece of knowledge she has in relation to the procedure they need to follow as *galiba bi dilekçe falan veriyoz* 'I guess we submit a letter or something' in line 8. As she is not completely sure about this procedure, she makes use of *falan* to mark her hesitation. Utilizing vague expressions appears as an inherent pragmatic strategy within the discourse units with the communicative purpose of *figuring things out* as the speakers jointly try to develop solutions at hand and they integrate estimations and suggestions rather than precise explanations while they are arriving at an understanding.

Joking Around

Though small in terms of the number of identified discourse units with *f(a)lan* (*n*=43), communicative purpose of *joking around [JOK]* covers various types of humourous interaction among young speakers of Turkish. Excerpt (57) below provides an example to the use of *f(a)lan* in such discourse units in the CoTY. The conversation is among speakers who are 16-year-old three male friends from İzmir. The speakers engage in collaborative humourous interaction which is initiated by SM11006's question *on sekize girince ne olacak sence?* 'what do you think will happen when we turn eighteen?' in turn 1 below.

(57) Y-3-M07122020-a

1	SM11006	((laughs)) ((name_SM11007)) on sekize girince ne olacak sence? ***((laughs)) ((name_SM11007)) what do you think will happen when we turn eighteen?***
2	SM11007	kanka işte üniversiteye <u>falan</u> gidince • böyle hep kızlar şey yapıyomuş. ***dude, when you go to college <u>and stuff</u>, they say it is the girls doing you know.***
3	SM11005	teklif ediyomuş. ***asking out.***
4	SM11006	teklif ediyomuş galiba. ***they are the ones asking out, I guess.***
5	SM11007	aynen. o yüzden çok heyecanlı bi durum. ***exactly. that why it is a very exciting thing.***
6	SM11006	gerçekten. ***really.***

SM11006's question in the first turn is accompanied by laughter which suggests that the question is posed at his friends to invite them for light-hearted talk. To respond to that call, SM11007 takes the turn and puts forward a topic which he assumes all the speakers in the conversation are familiar with. In turn 2, he begins the topic with the utterance *kanka işte üniversiteye falan gidince* 'dude when you go to college <u>and stuff</u>' in which general extender *f(a)lan* is used to trigger the shared conceptualizations other also speakers have regarding college life. The discourse unit displays that the speakers are able to establish the shared conceptualization regarding college life and love life as SM11005 continues SM11007's narrative in turn 3, followed by SM11006 swift alignment with the topic in turn 4. SM11007 approves the constructed narrative by responding with *aynen* 'exactly' which shows that the speakers created this discourse unit of *joking around* collaboratively and the general extender *falan* acted as the initiator for this particular episode of interaction.

Situation-dependent Commentary

The purpose of *situation-dependent commentary* (*n*=38) occurs in contexts in which the speakers are talking about entities, people, or event in their immediate situational context. The distribution of main and sub-topics previously indicated that the speakers in the CoTY often talk about the ongoing activities they perform while speaking to each other. An example is presented in (58) below in which two 16-year-old speakers from

Ankara are talking to each other via online communication channels. One of the speakers, SF10008, is skating and speaking at the same time.

(58) Y-3-F-14052021

1	SF10008	dün gece güveç yaptık da. onun kabını annem plastik… ***last night we baked a casserole. the pot we used for it, my mom (put it in) a plastic…***
2	SF10009	güveç? ***casserole?***
3	SF10008	hı? ***huh?***
4	SF10009	güveç mi? ***is it casserole?***
5	SF10008	güveç. ***casserole.***
6	SF10009	güveç. ***casserole.***
7	SF10008	evet güveç. güveç. **yes, *casserole. casserole.***
8	SF10009	tamam. ***okay.***
9	SF10008	onu yaptık da az önce onu ısıtmışlar. işte bi tane plastik şeyin içine koymuşlar. yamulmuş. boyanmış falan filan. ben onu atmaya gidiyorum. ***we did that, but they just heated up earlier. they put it in a plastic thing. it's wraped. it's smudged and stuff. I'm going to throw it out.***

In the interaction presented in (58), SF10008 tells her friend that she is skating outside, going to a rubbish bin on the street to throw out the rubbish from dinner. She starts to depict the distorted form of the pot to her friend in turn 1, and she explains the reasons behind the distortion of the shape she is throwing in the bin in turn 9. While explaining, she uses general extender *falan* in the utterance *yamulmuş, boyanmış falan filan* 'it's wraped. it's smudged and stuff' to mark the assumption that her friend knows the process of deformation of a pot in the extreme heat. As a result, she refrains from giving all the details and concludes the depiction of the immediate situational context by the utterance *onu atmaya gidiyorum* 'I'm going to throw it out'.

Describing or Explaining the Future

Future-oriented discourse units (*n*=35) include speakers' comments about the future, their plans and intentions, as well as their hypothetical visions for the future. In the CoTY, a salient topic within the discourse units with the communicative purpose of *describing*

or explaining the future [FUT] is dreams. As a form of a hypothetical vision for the future, dreams are co-constructed in a shared hypothetical space which is linguistically marked by *f(a)lan* among the young speakers of Turkish. The use of general extender *f(a)lan* in such cases assumes that the interlocutor shares the same conceptualizations for the future. In excerpt (59) below, for instance, 16-year-old speakers from Eskişehir are dreaming about going abroad together and one of the speakers makes use of *falan* to indicate that he is in the same conceptual territory as his friend.

(59) Y-3-2M1F-16052021-a

1	SM10006	şeyi hayal ediyorum. Danimarka'ya gittiğimi. veya Hollanda'ya gittiğimi • Amsterdam'a. ***I have this dream. that I'm going to Denmark. or Netherlands. Amsterdam.***
2	SM10005	Amsterdam'da şu an bir fotograf çekindiğimizi düşün! Alllah'ım şu an mutlu oluyorum! ama/ ***imagine that we are taking a photo in Amsterdam right now! God I'm feeling happy now! but/***
3	SM10006	((name_SM10005)) düşünsene Amsterdam'da gezdiğimizi! veya ((name_female))'in bizi Amsterdam'a ziyarete geldiğini! ((laughs)) ***((name_SM10005)) imagine that we are strolling around Amsterdam! or ((name_female)) visiting us in Amsterdam! ((laughs))***
4	SM10005	((laughs))
5	SM10006	ya gerçekten aşırı eğlenceli olmaz mı? beraber bisiklet sürerek falan geziyoruz! veya… ***wouldn't it be really fun? we are biking everywhere and stuff! or…***
6	SM10005	mükemmel olur! ***that would be perfect!***

In (59) above, SM10006 initiates an episode of co-construction of a dream through the utterance *şeyi hayal ediyorum* 'I have this dream' in turn 1 and the hypothetical future is jointly expanded through following turns by speakers. In turn 2, SM10005 visions an activity (i.e., taking photos) they could do if they would visit Amsterdam and in the following turn of 3, SM10006 proposes another activity they could enjoy doing together (i.e., strolling around the city). In the same turn, SM10006 expands storyline of the dream by getting a mutual friend of theirs involved in this the hypothetical narrative through his utterance *(düşünsene) ((name_female))'in bizi Amsterdam'da ziyarete geldiğini!* '(imagine) ((name_female)) visiting us in Amsterdam!'. Later in turn 5, SM10006 uses general extender *falan* in his utterance *beraber bisiklet sürerek falan geziyoruz!* 'we are biking everywhere and stuff!' to convey the message that he assumes that SM10005 would approve a new member to their group and would have similar plans as well. By

using *f(a)lan,* then, SM10006 intends to mark social cohesion and assert ingroupness with his friend. As a response, in turn 6, SM10005 confirms that he is in solidarity with his friend in the future-oriented conceptual space as well.

Giving Advice and Instructions

Similar to what Biber at al. (2021) reported for the BNC2014, discourse units which have the communicative purpose of *advice giving [ADV]* and contain *f(a)lan* are relatively infrequent in the CoTY overall. Though limited in terms of occurrences, the majority of *f(a)lan* tokens (*n*=9, observed in 60% of ADV discourse units) are used particularly in the offers, suggestions, or instructions regarding school work. Excerpt (60) below is an example for the discourse unit with this communicative purpose in which a 14-year-old speaker from Kırklareli is giving advice to her interlocutor regarding a homework by using *f(a)lan* multiple times.

(60) Y-2-F-02122020

1	SF09004	ya • şöyle söyliyim sana • ya evet ya internetten almak zorundasın. illa ki bakmak zorundasın. ama mesela hani copy paste yerine kitaptan • işte bizim kitapta var bu arada üç konuda past perfect de var. baktım şimdi. yüz seksen birde falan var bir de yüz yirmi beşte var. ***well, let me tell you this. yes, you have to copy from the internet. you have to check. but rather than doing copy and paste, from the book. we had the subject of past perfect tense (in English) in three topics. I have just checked. it is on the page hundred and eighty or something, and it is also on page hundred and twenty-five.***
2	SF09003	tamam bakarım. ***okay I'll check it.***
3	SF09004	ondan sonra oralardan falan bakıp ondan sonra internetten falan bakıp böyle yazabilirsin. ama şimdi vaktimiz de daraldı yani • bu hafta içinde atmamız lazım ki onu yapman senin sürer biraz. hani o yüzden bence direkt atabilirsin. ‿yani ama yine çalış konuya yani atıyorum sana bir soru sorduğunda sen öyle mal gibi kalma • ki simple'la continuous'u anlattı. ‿onları bil bence. ***then you can check from those pages and stuff and then look at those on the internet and stuff and write up. we have to submit it this week and it takes some time to do it. that's why I think you can just submit it. but I mean, study the subject. when he (teacher) asks you something, you don't get petrified like a dummy. he previously explained simple past and past continuous tense. I think you should know them.***

4 SF09003 onları bilmiyorum çünkü dersi dinlememişim büyük ihtimal. hatırlamadığıma göre ((laughs)).
I don't know those topics because I probably didn't listen to the lecture. I don't remember anything at all ((laughs)).

The advice given by SF09004 in this excerpt is about preparing a paper to submit to the teacher as homework. In turn 1, SF09004 provides a detailed explanation for the issues SF09003 needs to pay attention to. The first instance of *falan* present in this line *yüz seksen birde falan var* 'it is on the page hundred and eighty or something' marks an approximation with regard to the information given. In turn 3, two additional instances of *falan* occurs when SF09004 refers to the types of resources she previously suggested for her friend in turn 1. In a discourse unit with the communicative purpose of giving advice, *f(a)lan* displays organizational functions as in the case of excerpt (60).

Engaging in Conflict

The final communicative purpose is *engaging in conflict [CON]* which is reported among the least frequently observed communicative purpose type by Biber at al. (2021) for the BNC2014 data. Likewise, among the CoTY data containing intances of *f(a)lan*, the scope of this particular purpose was found to be relatively limited (*n*=9). This infrequency is most probably due to the inherent characteristic of the register of the corpus which is the informal talk among close friends. In this type of communicative purpose, the interaction is marked for the presence of disagreement, be it light-hearted teasing or a more verbally aggressive debate.

Below in (61), a discourse unit in which a conflictual talk between a speaker group of three friends from Çanakkale is presented.

(61) Y-3-2M1F-14052021

1 SM11004 gelmedin.
you didn't show up.
2 SM11010 kanki • annemler dedi. marketten bişeler alıncak dedi. gittim abi. telefonu da bıraktım o gün biliyo musun?
kanki, my mother came. she told me to buy some stuff from the süpermarket. I went there, bro. I didn't have my phone with me that day, you know.
3 SM11004 aradım.
I called you.

4	SM11010	normalde hiç bırakmam. ondan sonra. abi baktım siz yazmışsınız. yok kanka diyo günaydın. ((name_SF11008)) bana trip yapıyo. orda günaydın günaydın diyo. ((laughs)) ***I don't normally leave it. then. I saw your messages, bro. it says 'kanka good morning'. '((name_SF11008)) is sulking'. saying 'good morning, good morning'. ((laughs))***
5	SM11004	((laughs))
6	SM11010	((laughs))
7	SF11008	trip falan atmadım orda! ***I didn't sulk or anything!***

In the excerpt (61) above, 17-year-old speaker SM11004 is scolding their friend SM11010 for previously breaking his promise of meeting them in turns 1 and 3. In turn 4, SM11010 tries to defends himself by teasing the other participant in the conversation, SF11008, that she was overreacting for sulking at him just because he didn't show up. In turn 5, though, SF11008 responds with an exclamatory utterance of disagrement *trip falan atmadım orda!* 'I didn't sulk or anything!'. This utterance marks the point where conflict is linguistically manifested. Though it was directed at SM11010 as a response to his teasing aimed at her, SF11008 integrates general extender *f(a)lan* as a mitigator to soften the force of her utterance oriented at SM11010. The results indicate that in discourse units with the communicative purpose of engaging in conflict, then, *f(a)lan* is used to avoid the conflict rather than to engage in it.

Overall, the distribution of general extender *f(a)lan* across the communicative purposes shows that the pragmatic functions of vague language, in this case Turkish general extender *f(a)lan,* are influenced by their local contexts. The analysis indicates that there are functions of *f(a)lan* which are identified to be salient in particular types of communicative purposes. For instance, within discourse units which have the communicative purpose of sharing *personal feelings and evaluations [FEL],* general extender *f(a)lan* is particularly utilized as a hedging device in episodes of gossip talk. In the second most frequently identified communicative purpose after FEL, the communicative purpose of *describing or explaining the past [PAS]*, it is found that *f(a)lan* is used as a discursive device to construct the episodes of reenactment as well as as a mitigatory to protect potential threats to speaker's positive face.

The communicative purposes of *situation-dependent commentary [SDC], describing or explaning (time neutral) [DES], figuring-things out [FTO], giving advice and intructions*

[ADV] reflect the register characteristics of the corpus, as speakers are frequently engaged in activities in their immediate contex (i.e. SDC), talk about their daily routines (i.e., DES), and studying for their lessons and exams (i.e., FTO and ADV).

As the CoTY includes data of casual conversation among friends, the communicative purpose of *engaging in conflict [CON]* is naturally infrequent the corpus. Still, the analysis regarding the pragmatic uses of *f(a)lan* in CON indicated that it is used as a mitigator to avoid conflict rather than increasing the verbal aggression in interaction.

In discourse units with the communicative purpose of *describing or explaining the future [FUT],* the young speakers of CoTY utilize *f(a)lan* to jointly construct dreams and maintain solidarity in a future-oriented hypothetical space. Finally, in episodes of *joking around [JOK], f(a)lan* exhibits a similar function observed for FUT, and is used as a pragmatic device for inviting the interactants to collaboratively construct the humourous talk. What is noteworthy is that speakers make use of general extender *f(a)lan* across all nine distinct communicative purposes in the data.

To sum up, the results of this study echo the arguments of scholars (Overstreet, 1999; Cheshire, 2007) who emphasized that the pragmatic functions of vague language should be examined in their local context. Adding on to this, the study proposed a systematic approach to examine by adopting the taxonomy developed by Biber et al. (2021) and Egbert et al. (2021). The analysis based on this taxonomy confirmed that the immediate context and the salient communicative purposes of these context influence the pragmatic functions of vague expressions.

In the following section of this chapter, the final group of interactional markers intensifiers will be presented.

4.3.4 Intensifiers

Intensification is linguistically operationalized by various linguistic devices and strategies in order to exaggerate or diminish the message conveyed. To illustrate, in the examples below, the underlined lexical items work as intensifiers which boost the meaning of the lexical item(s) they modify in English:

(i) I greatly admire his paintings. (verb modifier)

(ii) The play was a terrible success. (noun modifier)

(iii) The article was extremely interesting. (adjective modifier)

(iv) He was driving very quickly. (adverb modifier)

(v) He is much in favour of the US attack on Afghanistan. (PP modifier)

Intensifiers are productive in the sense that they have capacity to emerge and spread in short periods of time as well as re-emerge in new syntactic and semantic contexts (Aijmer, 2020; Nevalainen & Rissanen, 2002; Tagliamonte, 2008). Due to their dynamic nature, intensifiers are dubbed as 'fashion-victims' by Blanco-Suárez (2010) as they can fall out of use when they are overused, diffused, or used long-term which leads to decrease in their expressive power of capturing attention or conveying novelty (Aijmer, 2018; Bolinger, 1972; Tagliamonte, 2008). Though an intensifier may lose its salience in language use over time, diachronic studies show that they can be reactivated at another point in time as they are prone to renewal and recycling (Stoffel, 1901). Tagliamonte (2008, p. 391) also points out that intensifiers are not created 'ex nihilo' but rather a word which once used as an intensifier remains in the linguistic repertoire and a speaker can recycle this intensifier sometime later. They are highly expressive and can be used to reduce social distance (Aijmer, 2020; Fuchs, 2017; Irwin, 2014; Palacios & Núñez, 2012), express stance (Athanasiadou, 2007; Barbieri, 2008), and emotions (Méndez-Naya, 2003; Núñez-Pertejo & Palacios-Martínez, 2014, 2018; Tagliamonte, 2008). Because of this, they are often associated with certain groups, among them is youth. In this section of the current chapter, the types and patterns of intensifiers will be presented and discussed for Turkish youth talk.

4.3.4.1 Defining intensifiers

The linguistic devices used for intensification have been so far called by various names, among which 'intensifiers' (Bolinger, 1972), 'degree words' (Quirk et al., 1985), 'intensive adverbs' (Stoffel, 1901), and 'amplifiers' (Biber et al., 1999) can be noted.

There are two main approaches to classify intensifiers. The first one is the traditional and restricted categorization which focuses solely on adverbs. One of the earliest and most comprehensive categorization belongs to Quirk et al. (1985) who classifies adverbs

identified as intensifiers into 'amplifiers' and 'downtoners' in English. In this taxonomy, amplifiers are divided into maximizers (e.g., *completely*) and boosters (e.g., *very*) while downtoners are divided into approximators (e.g., almost), compromisers (e.g., *more or less*), diminishers (e.g., *partly*), and minimizers (e.g., *hardly*). Quirk et al. (1985, p. 590) points that amplifiers "scale upwards from an assumed norm" while downtoners has a "lowering effect". Amplifiers are divided into maximizers which "denote the upper extreme of a scale" (e.g., *absolutely, entirely, completely*) and boosters which "denote a higher degree" (e.g., *so, very, really*) (Quirk et al., 1985, p. 590). The problem with this classification is that intensification is solely based on gradeability and the subtypes are only guiding without clear-cut divisions. This classification emphasizes the function of 'degree modification' (Biber et al., 2002; Bolinger, 1972; Stoffel, 1901).

Recent cross-linguistic works, on the other hand, take a discourse and pragmatics-oriented approach and highlight that intensification is an evaluative phenomenon. As a result, it is suggested that intensifiers can be used with 'non-gradable bases' such as prefixes, nouns and verbs (Napoli & Ravetto, 2017; Paradis, 2001, 2008). The latter approach assumes that intensifiers can operate at the clause level and thus has a wider scope for linguistic devices and strategies labelled as intensifiers. This study treats intensifiers within this discourse-pragmatics oriented approach (please see 4.3.4.3 for the scope of intensifiers focused in this study) to explore the interactional facet of intensifiers among dyadic and multi-party Turkish youth talk.

Overviewing the existing studies on intensifiers, Tagliamonte (2008, p. 362) underlines that intensifiers display characteristics of "versatility and colour, capacity for rapid change, and recycling of different forms". As a result, they qualify as potential linguistic indicators for tracking linguistic change. Within this line, the recent work on intensifiers utilize corpus methods heavily and focus on semantic change, grammaticalization process, competition and recycling of intensifiers, and their distribution across various speaker groups, language varieties and registers (Méndez-Naya, 2008, p. 213).

4.3.4.2 Brief overview of related work on intensifiers

Intensifiers or intensification in Turkish language has received limited scholarly attention so far. The existing work is quite prescriptive and referential in nature which aims to illustrate the standard grammar of the language. As a result, in the first part of

this section, description and scope of intensifiers in Turkish will be presented, and in the following section, corpus-based spoken discourse studies exploring youth language will be outlined.

4.3.4.2.1 Intensifiers in Turkish

As indicated earlier, the scope of intensification varies. As for Turkish, intensification can be conveyed through various linguistic layers. It can be realized by means of a variety of linguistic devices such as prefixation (e.g., *yepyeni kitap* 'a brand new book'), suffixation (e.g., *küçücük çoçuk* 'a very small kid', *güzelce kız* 'a cutish girl'), adjectives (e.g., *çok keyifli* 'very fun'), reduplications (e.g., *güzel güzel çiçekler* 'very beautiful flowers', *dere tepe dolaştım* 'wandering a lot'), pronouns (specifically reflexive pronoun *kendi* 'self'), adverbs (e.g., *büsbütün haksız biri* 'a totally wrongful person'), postpositions (e.g., *tezini bile bitirdi* 'she even completed her thesis'), connectives (e.g., makaleyi yazdı, *hem de kısa sürede.* 'she wrote the article, and what's more in a short time') and interjections (e.g., *aha orada!* 'whoa there it is!'), particles (e.g., *güzel mi güzel bir tatil* 'such a nice vacation') in Turkish (Banguoğlu, 2011; Göksel & Keslake, 2005; Korkmaz, 2003; Lewis, 2000; Üstüner, 2003). All of these instances are used to strengthen (or in the case of dimunitive suffixation, to downtone) a particular aspect of the meaning of the item.

The existing studies on intensifiers in Turkish are restricted to the description or categorization of intensifiers within standard grammar of Turkish (İpek, 2016; Karaağaç, 2013) along with a few studies on connectives and their intensification functions (Çelik, 1999; Karaşin 2008; Yüceol Özezen, 2013). Pragmatic functions of intensifiers in contemporary spoken Turkish has yet to be investigated.

4.3.4.2.2 Intensifiers in youth talk

As with other work on youth language, intensifiers have been extensively studied using youth corpora of the COLT, the COLAm, the CORMA and the patterns are often compared with data from the SCoSE, the BNC1994, the BNC2014, LCSWE, the MLE. There are also small specialized corpora constructed as individual projects which explore the intensifier use in youth talk. Studies often focus on comparison of intensifier use with adult speakers, the influence of gender, and (dis)appearence of various types of intensifiers over time. Studies underline that intensifiers are prone to decline with age (Barbieri, 2008; Núñez-Pertejo & Palacios-Martínez, 2018; Tao & Xiao, 2007). Using the

COLT and the SCoSE data, Palacios & Núñez (2012) showed that teenagers use different intensification strategies than adults. They frequently use *really*, followed by *so, very*. Additionally, taboo and swear words such as *bloody* and *fucking* are also identified as intensifiers which are not observed in adult talk.

Echoing Labov's (1985) note regarding *really* as one of the most frequent intensifiers in American English and British English (1999), Tagliamonte's (2006, 2008, 2016) extensive works on intensifiers in Toronto English Corpus revealed that the most frequent intensifier was *really*, followed by *very, so* and *pretty* in Canadian English. The results show that age is a factor which correlates with the frequency of intensifiers. Intensifier really is used most frequently among speakers who are between the ages 20 to 29. Very was used most frequently among speakers over 50, so and pretty are most frequent among 13 to 19-year-olds. In other successive studies, the most frequently used intensifier was reported as really in English youth talk (Bauer & Bauer, 2002; Beltrama & Staum-Casasanto, 2017; Hessner & Gawlitzek, 2017; Ito & Tagliamonte 2003; Lorenz, 2002). Also, as a comparative study of intensifiers in youth versus adult talk, Núñez-Pertejo and Palacios-Martínez (2014) focused on maximisers absolutely and totally in youth talk from the COLT and adult talk from the DCPSE. The results indicated that both intensifiers are more flexible than they are in adult talk, and they appear to take up new functions, such as emphatic and affirmative response items, in youth talk.

Among the earlier studies, Stenström et al. (2002) noted that intensifier well was used frequently among young speakers of English in 1990s and considered it as typical feature of London youth talk. The COLT data revealed gender difference with regard to use of well in the corpus; boys used it as an intensifier more frequently than girls did. Building on the observations on *well* in British English, Aijmer (2020) adopted a diachronic perspective to monitor *well* as an intensifier. Comparing data from the BNC1994 with the Spoken BNC2014, the study revealed that *well* showed an increase in frequency and it displayed new functions over time. Social factors of age, gender, and social class are identified as the parameters influencing the new functions of *well*. The analysis showed that in both corpora, *well* is used more extensively by young speakers. Among its other functions, the results indicate that young speakers of English use *well* with 'slangy adjectives or particles' to establish in-groupness.

Recent studies which adopt a wider scope for intensifiers revealed that taboo words are saliently used as intensifiers in youth talk (Palacios-Martínez & Núñez-Pertejo, 2012). Taking a cross-linguistic perspective, Palacios-Martínez and Núñez-Pertejo (2014) illustrated that expletives used as intensifiers in English had religious connotations while Spanish expletives had sexual connotations.

Roels et al.'s (2021) comparative investigation of intensifiers used by Spanish youth makes use of the COLAm which was compiled between 2003-2007 and the CORMA corpus which was compiled between 2016-2019. Analysing the intensifiers within the scope of language change, the researchers indicated that there is a tendency for using more intensifiers as time went by. The top five intensifiers remained the same but their frequencies changed. The results are noteworthy in the sense that contrary to general assumption that intensifiers change rapidly, the intensifier types did not show any attrition for Spanish youth talk over a decade.

Macaulay (2002, 2006) conducted a series of research on youth talk in Glasgow. In his 2002 study on same-sex interactions, he indicates that intensifier use shows a socially stratified pattern. In terms of pragmatic function, Macaulay (1995, 2002) argued that intensifiers are used to show attitude of speaker such as indicating approval or using them as pejorative devices. In his 2006 work, he focuses on the in-group exclusive intensifier *pure* used by working-class adolescents in Glasgow in order to explore the linguistic changes in progress based on spoken data from 1997, 2003, and 2004. The study discusses *pure* as an 'unusual intensifier' which was not reported previously. The results show that it is used as an amplifier and as a sign of group identification. It is also noteworthy that the analyses revealed that frequency of occurrence is lower in 2004 therefore the study suggests that the intensifier *pure* may disappear over time.

4.3.4.3 Findings: Intensifiers in the CoTY

In this section, the inclusion criteria for the intensifiers included in the study, the procedure to identify them in the corpus, their types and distribution, speakers using them, functions and identified patterns of the intensifiers in the CoTY will be presented.

4.3.3.3.1 Types, distribution, and speakers

This study focuses on adjectival and adverbial intensifiers along with taboo intensifiers which have not received any substantial scholarly attention in Turkish. As presented in 4.3.4.2.1, while standard grammars of Turkish do not include swear words as intensifiers, the literature underlines the expressive power they have. In order to identify the tokens of lexical intensifiers in the corpus, two complementary sources are used to generate the potential intensifiers in youth talk in Turkish.

Firstly, a list of lexical items previously reported to be used for degree modification in Turkish language (Banguoğlu, 2011; Göksel & Kerslake, 2005) and youth talk in other languages was compiled, and secondly emergent list of tokens identified during the corpus construction stage was integrated into the list. The final list of intensifier candidates yielded 33 lexical items. Queries were run for each of the candidate items using the EXAKT tool of EXMARaLDA. These queries retrieved 29 types of 5389 tokens as potential intensifiers or intensifier heads. Later, concordance lines and their expanded contexts were qualitatively investigated for these tokens. In line with the scope of intensifiers to be included in this study, false starts, incomplete utterances, stand-alone tokens in single turns, and tokens which function as nouns, interjections, and discourse markers were excluded from the analysis. This analysis identified 2856 tokens used for intensification in Turkish youth talk. Following Biber et al. (1999), this study adopts the binary categorization of 'amplifiers' which are used to intensify the strength of a particular aspect of the meaning of the item and 'downtoners' which function to reduce this effect.

There are 29 types of 2856 tokens of intensifiers in the corpus. Table 34 below lists the types of tokens under the main groups of amplifiers and downtoners and their frequencies tabulated by speakers in the corpus. The table shows the total number of tokens retrieved from the corpus (*TN*) for each type, the absolute frequencies of tokens identified as intensifiers (*AF*) along with their relative frequencies (*RF*) per million in descending order. For each type of intensifier, total number of unique speakers, number of female speakers and male speakers are also presented to illustrate the extent each intensifier is used by the speakers of the the CoTY.

Table 34 Amplifiers and downtoners tabulated by frequencies and speakers

Category	*Type*	*English gloss*	*No. of tokens*			*No. of speakers*		
			TF	*AF*	*RF*	*All*	*Female*	*Male*
amplifier	çok	*very*	2101	**1705**	10103.8	102	57	45
	bayağı	*excessively*	325	**188**	1114.09	55	38	17
	en	*the most*	323	**148**	877.05	59	34	25
	fazla	*excessively*	166	**120**	711.12	53	31	22
	gerçekten	*really*	263	**112**	663.71	40	30	10
	aşırı	*excessively*	111	**109**	645.93	33	23	10
	cidden	*seriously*	106	**48**	284.45	23	19	4
	gayet	*excessively*	41	**37**	219.26	17	10	7
	full	*full*	42	**33**	195.56	23	12	11
	kesinlikle	*absolutely*	41	**17**	100.74	15	12	3
	valla(hi)	*really*	86	**15**	88.89	14	8	6
	harbi(den)	*really*	48	**14**	82.96	13	3	10
	özellikle	*particularly*	25	**14**	82.96	12	8	4
	iyice	*quite*	21	**13**	77.04	12	9	3
	iyi	*well*	403	**12**	71.11	10	6	4
	ana +	*mother-plus swearing exp.*	54	**9**	53.33	5	1	4
	öyle	*so*	602	**8**	47.41	8	3	5
	manyak	*crazy*	26	**5**	29.63	3	2	1
	tamamen	*completely*	32	**5**	29.63	5	3	2
	süper	*super*	11	**3**	17.78	2	2	0
	am +	*vagina-plus expletive*	135	**2**	11.85	10	1	9
	deli	*lunatic*	17	**2**	11.85	2	0	2
	epey	*quite*	1	**1**	5.93	1	1	0
	müthiş	*awesome*	14	**1**	5.93	1	0	1
		Sub-total	*4994*	*2621*	*15532*	*97*	*54*	*43*
downtoner	biraz(cık)	*barely*	346	**196**	1161.5	72	47	25
	bir tık	*a bit*	29	**26**	154.08	12	9	3
	azcık	*slightly*	13	**10**	59.25	7	3	4
	hafif	*slightly*	6	**2**	11.85	2	1	1
	bir miktar	*a bit*	1	**1**	5.93	1	1	0
		Sub-total	*395*	*235*	*1392*	*103*	*58*	*45*
Total			*5389*	***2856***	*16871*	***113***	*58*	*55*

TN: Total number of tokens in corpus, *AF:* Absolute frequency, *RF*: Relative frequency per million

The results show that the most frequently occurring intensifier in corpus is *çok* 'very' (*AF*=1705, *RF*=10103.82) which also ranks as the 7th most frequent token in the whole corpus. It is noteworthy that this intensifier is used ten times more frequently than the second most frequently occurring intensifier *bayağı* 'excessively' (*AF*=188, *RF*=1114.09) in the corpus. Intensifier 'very' is also reported to be the most frequent amplifier in British and American English (Biber et al., 1999). To present this conventional amplifier, below is an example for *çok* 'very' from the corpus:

(62) Y-2-F-02122020

1	SF09003	bişey söylicem • fotoğrafı atar mısın çok merak ettim. fotoğraf nasıl bişey yani ne alaka? ***you know what, could you send me the photo? I am very curious. what kind of a photo is it? what's the relevance?***
2	SF09004	ya • bilmiyorum ama bence çok tatlı. bayağı tatlı bir fotoğraf bence. ‿çok hoşuma gitti. ne biliyim böyle bi fotoğraf. ***well, I don't know but I think it is very sweet. I think the photo is so sweet. I like it very much. that kind of a photo.***

In excerpt (62) above, 15-year-old female speaker SF09004 uses *çok* to intensify the degree of her liking for a photo. In turn 2, she first conveys her opinion about the photo by stating *bence çok tatlı* 'I think it is very sweet' in which intensifier *çok* 'very' is used to modify the adjective *tatlı* 'sweet'. Following this uterance, the speaker modifies the same lexical item with the second-most frequently used intensifier *bayağı* 'excessively, quite' and uses *çok* 'very' to modify a verb and express the intensity of affection she has towards the photo.

Though this study scrutizes the pragmatic and disursive dynamics of interactional markers in the corpus rather than specifically exploring the effects of social categories such as gender over these practices, the existing literature places the investigation of the relationship between gender and the intensifier use at the heart of the research. These studies argue that variation on linguistic practices is predicated on gender differentiation, and women are often associated with frequent intensifier use (Fuchs, 2017; Lakoff, 1975; Murphy, 2010; Ito & Tagliamonte, 2003; Tagliamonte, 2005, 2008; Tagliamonte & Roberts, 2005). Still, there are contrasting results even when two studies examine the effect of gender by focusing on the same list of intensifiers in a data of similar designs (i.e., the BNC1994 and the BNC2014). Hessner and Gawlitzek's (2017)

study, for instance, reported that there were no gender differences in intensifier use in the BNC2014 as opposed to the findings of Xiao and Tao (2007) who reported that women used more intensifiers than men did in the BNC1994. Nevertheless, the existing body of corpus-oriented research has not yet provided a consistent answer regarding gender related patterns and intensification. In this line, the following part of this section will provide an account of the relationship between intensifier use and sex of the speakers in the CoTY. For this purpose, after exploring the patterns young female and male speakers exhibit in terms of the types of intensifiers they use in the corpus, the frequencies for the types of intensifiers with regard to sex of the speakers and the types of speaker groups were examined by making use of statistical tests.

In the CoTY, the intensifiers are used by 92% of the all speakers in the corpus (n=113), among them 58 speakers are female and 55 of them are male. Though low in number and thus evaluated as idiosynctatic uses, intensifiers *deli* 'lunatic' and *müthiş* 'awesome' are identified to be exclusively used by young male speakers while intensifiers *süper* 'super', *epey* 'quite', and *bir miktar* 'a bit' are exclusively used by young female speakers in the corpus.

With regard to the distribution of data according to the intensifier groups, amplifiers are used by 97 speakers in the corpus of which 54 speakers are female and 43 speakers are male while downtoners are used by 103 speakers which consist of 58 female and 45 male speakers.

Table 35 below illustrates that the order of most frequently used intensifiers differs for females and males except for *çok* 'very' (ranks first); *fazla, aşırı* which both mean 'excessively' (ranking fifth and sixth, respectively); and full (which ranks ninth). Additionally, amplifier *cidden* 'seriously' and downtoner *bir tık* 'a bit' occurs only in the top ten list of female speakers while vagina-plus swearing formule *am+* as an amplifier is exclusive to the top ten intensifier list for male speakers. In female speakers' top ten, there are two downtoners *biraz(cık)* and *bir tık* while male speakers' list only includes *biraz(cık)*, and *bir tık* which ranks quite low with a ranking of sixteenth in the complete list.

Table 35 The most frequent 10 intensifiers for females and males

*Rank**	*Tokens by female speakers*				*Tokens by male speakers*			
	A/D	*Type*	*Gloss*	*AF*	*A/D*	*Type*	*Gloss*	*AF*
1	A	çok	*very*	1200	A	çok	*very*	505
2	D	biraz(cık)	*slightly*	144	A	bayağı	*quite*	73
3	A	bayağı	*excessively*	115	A	en	*the most*	66
4	A	gerçekten	*really*	90	D	biraz(cık)	*barely*	52
5	A	en, fazla	*the most, excessively*	82	A	fazla	*excessively*	38
6	A	aşırı	*excessively*	73	A	aşırı	*excessively*	36
7	A	cidden	*really*	42	A	gerçekten	*really*	22
8	D	bir tık	*a bit*	21	A	gayet	*excessively*	20
9	A	full	*full*	18	A	full	*full*	15
10	A	gayet	*excessively*	17	A	am +	*vagina-plus swear*	12

A: amplifier, D: downtoner, AF: absolute frequency

**intensifiers with same AFs ranked together*

Total number of intensifier tokens (*n*=1955) produced by female speakers is twice as much as that of male speakers (*n*=901) in the corpus. In order to examine whether there is a significant difference between the frequencies for types of intensifiers used by female and male speakers in the corpus, a chi-square test was administered[40] (see Appendix J details on the results). The results revealed that with a single exception of amplifier *fazla* 'excessively', the difference in terms of intensifier frequency is significant for the rest of the intensifiers. The results statistically show that intensifiers mother-plus swearing expressions, *gayet* 'excessively', *harbi(den)* 'really', and *öyle* 'so' are used more frequently by male speakers in the corpus all of the remaining intensifiers are more frequently used by female speakers. In this sense, the results corroborated the previously reported results by studies (Fuchs, 2017; Lakoff, 1975; Murphy, 2010; Precht, 2008; Ito & Tagliamonte, 2003; Tagliamonte, 2005, 2008; Tagliamonte & Roberts, 2005) that female speakers favour intensifiers more than males.

An additional analysis was conducted to test whether speaker groups (all-female, all-male, mixed-speaker groups) showed any significant difference in terms of the number of intensifiers they used in their speech. Results of the chi-squared test showed that

[40] Among identified 29 types of tokens, chi-square test was only administered to most frequently occurring 20 intensifiers in accordance with chi-square test assumption that observed value of for each category should be greater than 5. The analysis was conducted in SPSS Statistics 28.0.1.

(with Bonferroni correction p<0,01666) except for amplifiers *bayağı* 'quite, excessively', *en* 'the most', *gayet* 'excessively', *full, vallahi* 'really', *harbiden* 'really', *öyle* 'so', *manyak* 'crazy', all intensifiers are used more frequently in all-female groups. Concerning mixed speaker groups, amplifiers *en* 'the most', *fazla* 'excessively', *gayet* 'excessively', *gerçekten* 'really', *özellikle* 'particularly' and downtoner *bir tık* 'a bit' are used significantly more frequent in mixed speaker groups compared to all-female groups and all-male groups in the corpus. Additionally, the results show that in mixed speaker group data, intensifiers *full, vallahi* 'really', *öyle* 'so', *manyak* 'crazy' are used significantly more frequently compared to all-female groups and intensifiers *çok* 'very', *kesinlikle* 'absolutely', *iyice* 'quite' are used significantly more frequently compared to all-male groups. Overall, the analysis indicates that gender of the speakers in a group influence the frequency of specific intensifiers used in interaction among young speakers of Turkish. Results regarding the pragmatic uses of amplifiers and downtoners will be presented in more detail in the following sections.

4.3.4.3.2 Amplifiers

The group of amplifiers show more variety as this group is made up of 24 types of intensifiers of 2621 tokens while downtoners is a smaller group of intensifiers which consist of 4 types of 225 intensifiers in total. Intensifier *çok* 'very' is the most frequently used amplifier followed by *bayağı* 'excessively' (*AF*=188, *RF*=1114.09) and *en* 'the most' (*AF*=148, *RF*=877.05).

Amplifiers include lexical items and phrases from the domain of taboo and swear words in Turkish which is not observed for general spoken Turkish represented in the STC. This CoTY-specific group of intensifiers include mother-plus swearing expressions *ana+* (*AF*=9, *RF*=53.33), vagina-plus swearing expressions *am+* (*AF*=2, *RF*=11.85), as well as other swear words of *manyak* 'crazy' (*AF*=5, *RF*=29.63), and *deli* 'lunatic' (*AF*=2, *RF*=11.85).

In excerpt (63) below, swearing expressions in turn 1 and 8 are used as amplifiers by 17-year-old male speakers SM11001 and SM12002, respectively. In this conversation, the speakers are talking about SM11001's ongoing platonic love for a girl. SM11001 shares with his friend that he struggles to carry on the conversation with the girl as she does not expand on his comments or rarely replies back to her messages.

(63) Y-2-M-19112020-b

1	SM11001	yazmıyorum anasını. ***I will not fucking text.***
2	SM11002	yaz ya! ***text her!***
3	SM11001	ya aslanım/ (vocative: my lion) ***well aslanım/***
4	SM11002	bak Barış Manço'nun bi lafı var. yaz dostum! ***look Barış Manço has this saying: 'write, my friend!'***
5	SM11001	bence zaten konuşma bitmiştir. ***I think the talk is already over.***
6	SM11002	bittiği gün bitmiştir. ***it is over when it is over.***
7	SM11001	ya bence bitmiştir. bak bence ben bi daha yazmazsam yazmicak. ***well, I think it is over. look, I feel that if I don't text her, she will not text me.***
8	SM11002	aslanım çünkü ayıp ediyon amına koyim kıza. oğlum • peşinde koşuyosun bi yıldır. yani şimdi bırakırsan senin kafana sıçayım. ***aslanım, because you are behaving fucking disgraceful to the girl. dude, you have been going after her for a year. if you give it up on this now, fuck you.***

In line 1, SM11001 is using a swearing expression to convey his disappointment about the situation and that he gave up on trying to get close to her. The swear word *anasını* which roughly corresponds to English 'fucking' highlights the anger and disappointment the speaker experiences. In the following turns of 2, 4, and 6; SM11001's interlocutor SM11002 tries to convince his friend to continue writing to the girl. Finally in turn 8, there is another instance of a swearin formule *am+* 'vagina-plus swear word' which again acts as 'fucking' in English. In this second instance, SM11002 uses the swearing expression to convey his opinion in a more intensified manner and with the purpose of changing SM11001's mind.

Having presented the swear words used as intensifiers in the corpus, another group of lexical items which are used for intensification are the loan words. As far as loan words are concerned, the analysis identified *full* (*AF*=33, *RF*=195.56) which is a borrowing from English and *süper* (*AF*=3, *RF*=17.78) which is an established anglicism of English 'super'. Intensifier *full* is exemplified in excerpt (64) below. In this conversation, 17-year-old female speakers from Denizli are talking about their German exam at school. SF12010 specifically refers to a question in the exam and states that she provided a made-up

answer for that question. To highlight the unexpected high mark she got from the exam, she inserts the intensifier *full* and intensifies that she 'completely' made up the answer.

(64) Y-3-F-06122020

1	SF12010	şey Almanca yazılısındaki ilk etkinliği hatırlıyo musunuz? ***well, do you remember the first task in German exam?***
2	SF12011	neyi? ***what?***
3	SF12010	ilk etkinliği. Almanca yazılısındaki. ilk soruyu. ***the first task. in German exam. the first question.***
4	SF12011	şey biz farklıydık. ***well, we got a different order of questions.***
5	SF12010	kedili bi soru vardı. ***there was a question with a cat.***
6	SF12012	hı-hı˙ ***mm-hmm.***
7	SF12010	işte ben orda var ya <u>full</u> sallamasyon yaptım. buna rağmen yetmiş beş almışım. yine iyi bence. ***at that part I <u>completely</u> made it up. still I got a seventy-five. not bad.***

As shown in (62), (63), and (64), young speakers of Turkish use amplifiers to highlight emotion-laden messages. Amplifiers are also used to underline personal opinions and thus express stance as exemplified in excerpt (65) below in which a 16-year-old male speaker from İzmir uses amplifier *gerçekten* 'really' (*AF*=112, *RF*=663.71) which literally means 'for real' to strengthen the force of his personal opinion regarding an absent other. The topic of the conversation is behaviours of a mutual friend with whom SM10004 is not on good terms. Throughout the conversation, SM10004 lists the types of behaviours he does not approve or like such as turn 1. To his dismay though, his interlocutor SF11006 does not judge the mentioned person on negative terms but rather states that she feels pity for this person in turns 2 and 4.

(65) Y-2-FM-0412220

1	SM10004	şey diyo işte • şapkalı kedi geliyo falan. böyle saçma saçma şeyler söylüyo. ***he says 'cat with a hat is coming' and stuff. he is telling these kinds of nonsense stuff.***
2	SF11006	abi! üzüldüm ama şu an! ***dude! I am feeling sorry now!***
3	SM10004	neyine üzüldün tam olarak? ***you are sorry for what?***
4	SF11006	bilmiyorum! üzüldüm şu an! ((laughs)) ***I don't know! I feel sorry! ((laughs))***

5	SM10004	ben de şey oluyorum arada • hehe ((imitating laughter)) yapıyorum azıcık. onu da yapmıyorum artık. gerçekten çok sahte çünkü. ***I sometimes go 'haha' at him a bit. I can't do that anymore. because (he is) really fake.***
6	SF11006	of! ***ugh!***
7	SM10004	komik değil. ‿bayağı şey oluyorum. Allah kahretmesin. bunu da yapmazsın. ***it is not funny. I go like 'God damn it. you don't do that!'***

In turn 5 in excerpt above, SM10004 shows that he acknowledges the tolerance SF11006 displays towards the said person and he implies that he used to be tolerant as well. SM10004 is firm about his opinion regarding that person and thus he uses the intensifier *gerçekten* 'really' to convince his interlocutor as well. In turn 6, SF11006 responds with engagement token *of!* to convey the message to SM10004 that she supports SM1004's assessment of that person.

Also in (65), notice that in turn 5, SM10004 uses downtoner *azıcık* 'slightly, a bit' to soften the evaluative force in his quotative utterance *arada hehe yapıyorum azıcık* 'I sometimes go haha a bit'. In this context, *azıcık* (*AF*=10, *RF*=59.25) mitigates the evaluative force of a potential imposition to his negative face. This case highlights the prominent pragmatic function of mitigation for downtoners which will be exemplified in detail in the following section.

4.3.4.3.3 Downtoners

As previously presented in Table 34, the most frequently occurring downtoner is *biraz(cık)* 'barely' (*AF*=196, *RF*=1161.50) followed by *bir tık* 'a bit' (*AF*=26, *RF*=154.08) and azcık (*AF*=10, *RF*=59.25) in the corpus. Other identified downtoners in the corpus; *hafif* 'slightly' and *bir miktar* 'a bit' are low in frequency (*AF*=2, *RF*=11.85 and *AF*=1, *RF*=5.93, respectively) thus can be considered as examples of idiosyncratic uses.

The analysis indicates that downtoners are used for pragmatic mitigation among the young speakers of Turkish in the corpus. For instance in (66) below, downtoner *biraz(cık)* 'barely' (*AF*=196, *RF*=1161.50) in the corpus displays a hedging function in interaction. The conversation between two 16-year-old female friends from Ankara are talking about a person they recently met:

(66) Y-2-F-14052021-5

1	SF10013	kız birazcık ((name_female)) vibe'ı veriyodu bayağı. di mi?
		(English)
		the girl was giving off a bit of a ((name_female)) vibe, a lot. wasn't she?
2	SF10014	yani.
		well.
3	SF10013	yani bilmiyorum. ben çok ((name_female)) havası aldım kızdan.
		well, I don't know. the girl reminded me of ((name_female)) a lot.

In the excerpt, it is noteworthy that SF10013 simultaneously uses an amplifier (*bayağı* 'excessively') and a downtoner (*birazcık* 'a bit') for the same message conveyed in a single utterance. It is an evaluative talk in which SF10013 is gossiping about both a girl they recently met and a common friend they knew. SF10013 firstly uses downtoner *birazcık* as a mitigator to refrain from face-threatening act of gossiping (Blum-Kulka, 2000) but also immediately reinforces her evaluative stance with the amplifier *bayağı* within the same utterance. SF10014 does not fully align with SF10013's negative evaluation of absent third parties as she responds merely with *yani* 'well' without expanding on the topic. In turn 3, SF10013 repeats her opinion, this time by utilizing amplifier *çok* 'very' in order to establish her stance regarding the girl.

Similarly in (67) below, a mixed group of 16-year-olds from Eskişehir talk evaluatively about their teachers. In turn 1, SM10002 states his opinion of one of the teacher's lecture style and teacher identity. He intends to criticize the harsh or strict behaviour the teacher imposes on them but expresses this observation by hedging it with downtoner *birazcık* 'a bit'.

(67) Y-3-2M1F-09052021

1	SM10002	((laughs)) yani dersi güzel anlatıyo. evet anlıyorum. ama ne biliyim. bazen birazcık sert olabiliyo yani. ((short laugh)) ama bu uzaktan eğitim döneminde ben de gerçekten <çok beğendim. >/1>.
		((laughs)) well he lectures well. yes, I comprehend the topic. but I don't know. well, sometimes he can be a bit harsh. ((short laugh)) but <I really liked >/1> him during this distance education period.
2	SF10016	<şimdi şöyle… >/2>
		<the thing is…>/2>

3	SM10002	((name_male)) hocadan. ***from ((name_male)) teacher.***
4	SF10016	((name_male)) hoca mükemmel bi insan. bi de bi tık size yurtta daha çok haşır neşir ya • o yüzden erkeklere karşı bi tık daha sert. ***teacher ((name_male)) is such a great person. and you know he is dealing with you in dormitory a bit, that's why he is a bit harsher towards the boys.***
5	SM10001	aynen. ***exactly.***
6	SF10016	bunu kabul edebilirim. ***I aggree with this point.***
7	SM10002	((laughs)) bi tık mı? ***((laughs)) a bit?***
8	SM10001	((short laugh))
9	SF10016	bi tık! ***a bit!***

In turn 4 in (67) above, female speaker SF10016 underlines that she has a high opinion of the teacher by depicting him as a *mükemmel bi insan* 'a perfect person' and justifies that the reason behind his strict behaviour could be due to his supervising duties in the dormitory. Similar to the mitigating use of *birazcık;* downtoner *bi tık* 'a bit', which is the second most frequent downtoner in the corpus (*AF*=26, *RF*=154.08), is used to soften the criticism posed at the teacher in the utterance *bi tık size yurtta daha haşır neşir ya, o yüzden erkeklere karşı bi tık daha sert* 'he is dealing with you in dormitory a bit, that's why he is a bit harsher towards the boys'. To this, SM10001 responds with convergence token aynen 'exactly' in turn 5 (See 4.3.1.3.4 for a detailed discussion on the response token *aynen*). In turn 7, SM10002 playfully asks *bi tık mı?* 'a bit?' (which is followed by SM10001's laughter) which further reveals that downtoner *birazcık* in turn 1 was used as a politeness strategy to refrain from a potential face threat oriented towards an absent other as well as speaker's positive face.

The multiple uses of *bir tık* in a single excerpt as presented in (67) led the researcher to scrutize this intensifier from a diachronic perspective. Thus in the following section, a more detailed account of downtoner *bir tık* will be presented.

4.3.4.3.4 Tracing delexicalization: From *tık* to *bi tık*

Delexicalization is a subprocess of grammaticalization which refers to the process of linguistic changes a lexical item undergoes which can be realized in various single or multiple levels such as phonetic reduction, decategorization, semantic change and

pragmatic shift (Bybee, 2003; Macaulay, 2006; Partington, 1993; Sinclair, 1992). In case of intensifiers, delexicalization is often observed when a lexical item partly or fully loses its original meaning and it is turned into an intensification marker (Tagliamonte & Roberts, 2005). As delexicalization is a continuum, a lexical item can be identified as fully or partially delexicalized. Partington (1993) defines this process as the modal-to-intensifier shift and indicates that it can be observed through both synchronic and diachronic linguistic evidence.

While currently there is no diachronic corpora for spoken Turkish, there are the TNC and the STC which provide snapshots of spoken Turkish from different periods of time. 50-million-word corpus the TNC consists of spoken and written data from 1990-2013 while 350,000-word specialized corpus STC is made up of entirely spoken data compiled between 2008-2013 in Turkey. For the purposes of tracing corpus evidence for the grammaticalization of intensifiers in spoken Turkish, two-word cluster *bir tık* 'a bit' which in majority acts as an amplifier in the CoTY will be the examined in detail via data from the TNC, the STC, and the CoTY combined. Following *aşırı* 'excessively', *bir tık* 'a bit' is the lexical item which functions as an intensifier more frequently (90% of the tokens function as intensifier) than all other lexical items with intensifying functions in the corpus. Sample concordance lines for *bir tık* is presented in Figure 16 below.

Left Context	Match ▲	Right Context
nın annesi eğitim konusunda katı. yani bi	tık	katı evet. çünkü yani derslerine çalışsın. etsin i
amin. ya abi ben bilmiyorum bi de benim bir	tık	şeyim var biliyorsun • mükemmelliyetçiliğim var.
layamayacağım gibisinden bakıyosun mesela. sen bir	tık	• bi konuda kitlenip kalıyosun açıkçası • bana gör
bi	tık	daha iyi dedim!
ceki bölümü daha çok sevdim. son bölüm de böyle bi	tık	güldürmedi. ‿yani evet şey anlamlı şeyler vardı.
ama bi	tık	güldürmedi yani.
cam ama • beli halledilebilir gibi geliyo. yani bi	tık	daralttırabilirim belki. ama arkasında da logosu v
öyle. sevgilisi varmış. ‿buna bi	tık	üzüldüm.
inin aynısını bulamıyorum piyasada. benimkinin bir	tık	üstü var da. aynısını bulmam lazım.
Endemik'e el sallayıp hemen ıı ˙ bi	tık	altına iniyoruz.
o yüzden bi	tık	daha matematiğe ağırlık veriyorum.
ırı ayrıntılı güvenmesem de tarihle coğrafyamın bi	tık	iyi olduğunu düşünüyorum. yani ordan da bi özgüven
ç kişiydi. ama hani bizim odadakilerle değil de bi	tık	daha yan odadakilerle falan böyle dokuz kız • topl
bi de bende şey başladı. hani kitap olunca bi	tık	elin yukarda kalıyo ya?
ya bak şöyle düşün. üniversitenin ilk senesi • bi	tık	daha kolay. hani tüm üniversiteler için demiyorum

Figure 16 Sample concordance lines for bir tık in the CoTY

Downtoners *bir miktar* and *bir tık* both roughly correspond to English downtoner 'a bit'. While *bir miktar* is an established degree modifier in Turkish, it only has a single occurrence in the corpus possibly due to its formal nature. Downtoner *bir tık,* on the other hand, occurs 29 times in the corpus and can be considered part of contemporary

Turkish slang. Typically, it reduces the degree or the intensity of an act or item it modifies as in (68) where a 16-year-old is telling her friend that she intends to fix the fitting of the trousers she plans to order online.

(68) Y-2-F-14052021-2b

1	SF09007	ee˙ mavi olan var ya • böyle açık renk. nerdeyse beyaz. onu çok beğendim. ***you know that blue one. the lighter shade. almost white. I like that one a lot.***
2	SF11011	evet. ***yes.***
3	SF11011	hı-hı˙ ***mm-hmm.***
4	SF09007	ve yani olmazsa iade ederim diye düşünüyorum. bunu da alabilirim. çünkü yirmi sekiz • bedeni yok. yani otuz almak zorunda kalcam ama • beli halledilebilir gibi geliyo. yani bi tık daralttırabilirim belki. ama arkasında da logosu var. ***if it does not fit I can return it. I can buy this one as well. because size twenty-eight is out of stock. I will have to buy size thirty but I think the fitting of the waist can be fixed. I mean, maybe I can get it narrowed a bit. but it has a logo on the back.***

Similarly in excerpt (69) below, a speaker uses *bir tık* as a downtoner and provides a meta-comment to clarify the meaning conveyed by this lexical expression. It is from an online conversation and 16-year-old speakers from Mersin who are talking about fasting in Ramadan[41]. In turn 6, SF10016 states that it is a bit difficult to fast and catch up with school work by *yani ben bi tık zorlandım* 'it was a bit difficult for me' followed by a roughly synonymous expression *azcık* 'a little' to further emphasize the meaning of *bi tık*.

(69) Y-3-2M1F-09052021

1	SF10016	oruç tutuyo musunuz siz? ***are you guys fasting?***
2	SM10002	evet! ***yes!***
3	SM10001	evet. ***yes.***
4	SF10016	Allah kabul etsin! ***May God accept!***

[41] In Islam, Ramadan is a one-month period of time in which Muslims practice a selection of religious practices. Among them is fasting which requires abstinence from food or drink from dawn to sunset.

5	SM10002	sağol! sağol! ***thanks! thanks!***
6	SF10016	Ramazan'da okul nası geçiyo? yani ben bi tık zorlandım. azcık ama. ***how is it like going to school during Ramadan? I mean it was a bit difficult for me. a little bit, though.***

The analysis indicates that almost half of the time (46%, *n*=12) downtoner *bi tık* also has the function of mitigation in facework. Reducing the expressed intensity of an act, speakers attenuate the illocutionary force of the utterance so that the established harmonious relationship among the interactants is not disrupted. An example to this use is excerpt (70) in which two 18-year-old friends talk about their current performances regarding their studies for national university exam. One of the speakers uses *bi tık* to express that she is 'a bit' good at the subjects of history and geography and this gives her confidence for the upcoming exam. In this case, downtoner 'a bit' is used as a hedge to a self-praise, because as a form of self-compliment, it is a potentially face threatening act (Brown & Levinson, 1987) for the speaker themselves as the utterance bears the risk that the speaker can be judged as pompous by the hearer (Pomerantz, 1978; Speer, 2012).

(70) Y-2-F-05122020-1

1	SF12008	bi de bişe diyim mi • benim sadece edebiyat değil TM'de • hani tarih coğrafya falan da var ya hani • benim tarihim hani • ayrıntılı olarak güvenmesem de • çok aşırı ayrıntılı güvenmesem de tarihle coğrafyamın bi tık iyi olduğunu düşünüyorum. yani ordan da bi özgüven geliyor bana. ***and you know what? not only the Literature, there are also History, Geography and stuff. though I am not fully confident, I think I am a bit good at History and Geography. I feel confident because of that.***
2	SF12009	valla o büyük artı ya! ***well that's a big advantage!***

Token *tık* occurs three times more frequently (*AF*=29, *RF*=171.85) in the CoTY compared to the STC (*AF*=12, *RF*=57.50). In the CoTY, 26 out of 28 tokens occur as the cluster *bir tık* while this cluster is not present in the STC at all. A separate KWIC analysis was conducted to investigate whether *tık* functions as an intensifier in any form in the STC. The analysis illustrated that *tık* is used as inanimate imitative[42] (Oswalt, 1994) which is

[42] sound produced by an inanimate item.

a form of onomatopoeia and it is present in single standing or reduplicated forms among adult speakers of Turkish as in (i) and (ii) below:

(i) VOL000447: adam orda sana tık tık tık hazırlıyor her şeyi.
'he tık tık tık prepares everything for you there.'
[source: STC-Beta, 024_100501_00160]

(ii) ATA000156: ki bin devirden z/ sonra tık ((0.1)) diyor.
'after a thousand rotations, it makes tık.'
[source: STC-Beta, 102_091223_00057]

In (i) reduplicated *tık* roughly corresponds to the meaning that the doer of the action accomplishes an action in a quick and orderly fashion. In this sense, *tık tık tık* can be idiomatically translated as 'quickly' while in (ii) it is used in its original imitative meaning to express the sound a mechanical device makes. Among these two uses, there is only a single use of *tık* which exhibits delexicalization and is used by a single speaker in (iii) with the meaning 'at all'.

(iii) VED000860: ((0.2)) tık arıza vermedi. ‿biliyor musun?
'it did not break down at all, you know?'
[source: STC-Beta, 073_100201_00338]

In contrast, speakers in the CoTY extensively use two-word intensifier cluster *bir tık* in which *tık* is a noun modified by indefinite article *bir* 'a/an'. Single standing form observed in (iii) occurs only once as *tık cevap yok* 'no answer at all' in the CoTY. The remaining uses are entirely downtoners as presented in (68), (69), and (70).

The STC was compiled between the years 2003-2007 and even though it provides limited data for the speech of younger speakers, this observation suggests that lexical item *tık* may be undergoing the process of grammaticalization. To expand the scope of analysis and data, the TNC data was explored. In spoken part of the TNC which has 1,000,000 words compiled between data from 1990-2013, *tık* occurs 35 times (*RF*=34.52) and all of them correspond to the uses presented in (i) and (ii) and no instance of *tık* as an intensifier was found. When the query was run for the written component of the TNC, the KWIC analysis of the retrieved tokens (*AF*=475, *RF*=9.56) yielded results which are

complementary to the aforementioned observations regarding *bir tık* as an emergent intensifier in contemporary Turkish spoken by younger speakers. The analysis shows that *bir tık* occurs four times as an intensifier in written portion of the TNC, but specifically in four separate blog posts by two authors published in 2012 and 2013. Though there is no metadata regarding the age of the authors in the corpus, the names of the blogs are provided. One of the blogs has a fashion and celebrity-gossip oriented content while the other one has posts about books, shows, places the author recommends. Public information regarding the profile of one of the blogs reveals that the fashion blog has a female author who was in her early twenties when she published the posts which has *bir tık* as intensifier (*n*=3) as in (iv) below in which the author criticizes the outfit of a celebrity in a blog post published in 2013.

(iv) üzerindeki büstiyerin bi tık daha uzun olmasını tercih ederdim.
'I would have liked if the bustier she had was a bit longer.'
[source: TNC-V.03, W-ZI45E1C-5072-232]

Though limited in terms of occurrence, what is noteworthy for these instances is that the authors adopted a conversational style in their writing containing words and expressions which are part of spoken Turkish. They write to address an audience within their virtual private domain. As a result, blog posts reflect the informal register of Turkish. The findings of the TNC corroborate the argument that *tık* may have undergone the process of delexicalization and transformed into *bir tık* as an intensifier over the last decade. The corpus evidence for the use of *bir tık* as an intensifier is traced back to 2012 in the language of social media used by potentially younger users of Turkish. Approximately ten years later, it is salient as a downtoner in the language spoken by Turkish youth in the CoTY.

4.3.4.3.5 Accentuating the personality traits: *aşırı* or *bayağı*

In the CoTY, intensifiers *aşırı, bayağı, fazla,* and *gayet* all semantically correspond to English intensifier *excessively*. Among them, *aşırı* (*AF*=109, *RF*=645.93) and *bayağı* (*AF*=188, *RF*=1114.09) stand out within the keywords in Turkish youth talk (see section 4.3). This section will compare identified patterns and functions of these two intensifiers in the corpus.

In order to observe their associations with other lexical items in their local contexts, firstly collocation analyses were carried out for both intensifiers. The analysis indicated that both *aşırı* and *bayağı* had adjectival *iyi* 'good' as their strongest collocate in the corpus (both had MI3 score above 10 and a T score above 2, occurred at least 5 times with the node word). After a close reading of expanded concordance lines for co-occurrences of *aşırı iyi (freq*.=8*, MI3*=10.825) and *bayağı iyi (freq*.=19, *MI3*=12.953), the results showed that both collocations are often used to express opinion about a person (i.e., self, each other, or an absent other) in the corpus (see Appendix K for the coded concordance lines).

This observation led to another layer of concordance analysis which focused on exploring the objects of intensification and identifying which traits or behaviours of people are the foci of intensification. The analysis showed that amplifier *aşırı* is used to emphasize negative traits of a person (*n*=11, 10%) while this use is limited for *bayağı* (*n*=5, 3%). On the contrary, *bayağı* is used to accentuate positive traits of a person more than it is used for negative-othering (*n*=22 and *n*=5, respectively) in the data. Excerpts (71) and (72) are typical examples for these uses. In (71), speakers are 17-year-old female speakers from İzmir and the topic of the conversation is a boy SF12006 once was in good terms with, yet not anymore.

(71) Y-2-F06122020

1	SF12007	abi birden herkesten uzaklaştı ama farkındaysan böyle. bütün dünyadan kendini soyutladı sanki. ***dude, you might have noticed that he alienated himself from everyone. it is as if he detached himself from the whole world.***
2	SF12006	evet evet evet. sanki böyle şey gibi davranıyo herkese • herkes benim düşmanım artık. bi/ sadece benim için işte • artık sadece kız önemli falan. ***yes yes yes. he is behaving like 'everyone is my enemy now. only the girl matters to me now' and stuff.***
3	SF12007	toksik davranıyo. aşırı toksik davranıyo. ***he is being toxic. he is being excessively toxic.***
4	SF12006	evet. ve toksikliğe karşı bişey yapınca da böyle bi düşünmüyo. ben yanlış mı davranıyorum diye düşünmüyo. direkt • şey yapıyo. ne denir? hani • ters tepip laf sokuyo geri. pasif. tamamen pasif agresif. ***yes. and when someone resists that toxic behaviour, he doesn't reflect on his behaviours. he is just doing, what is it called? it is like he backlashes and makes mean comments. passive. totally passive-aggressive.***

In turns 1 and 2, speakers complain about the behaviours of that person, and in turn 3, SF12007 depicts the behaviour of the said person as toxic. She first utters *toksik davranıyo* 'he is being toxic' and then immediately repeats her message with inserting the intensifier *aşırı* to further emphasize the unfavoured trait ***aşırı*** *toksik davranıyo* 'he is being excessively toxic'.

Excerpt (72) below is from a conversation among two 18-year-old female speakers residing in Istanbul. Speakers are graduates of high school and studying for university entrance exams. In the conversation which was conducted online, they are reminiscing about their high school years and talking about their mutual friends. SF13002 recalls that one of their friends had a very good command of English, and she uses intensifer *bayağı* to accentuate the high level of language proficiency the person had.

(72) Y-2-F-13122020

SF13002 hani bi de ((name_female))'nin İngilizcesi • bişe diyim mi bayağı iyi. bizden de çok iyiydi. hele lisedeyken • işte şey yapıyorken • ne yapıyorduk? biz bişe yapmıyoken ((short laugh)) ((name_female)) gelip teneffüslerde falan İngilizce kelimeler ezberliyodu.
by the way, ((name_female))'s English, let me tell you, it is quite good. she was far better than us. especially in high school, while doing, what were we doing? while we were not doing anything ((short laugh)), ((female_name)) would be memorizing vocabulary during breaks.

To contrast their functions, the representative cases presented in (71) and (72) suggest that aşırı marks a negative prosody while bayağı exhibits a relatively more positive prosody in discourse.

Additionally, it is observed that intensifier *aşırı* most of the time occurs in contexts with emotive involvement of the speaker (*n*=40, 37%) such as *aşırı bi şok oldum* 'I was quite shocked', *aşırı sinirlendim/duygulandım* 'I was quite angry/moved', *aşırı mutsuzum/seviyorum/hoşuma gidiyo* 'I am quite sad/I love it/I like it' while emotive involvement makes up only %11 of uses of *bayağı*. The majority of targets of intensification *bayağı* orients are aspects of concepts, objects, experiences and actions (*n*=88, %43) such as the uses of *bayağı* in *bayağı kapsamlı bir siteydi* 'it was a quite comprehensive website', *bayağı büyük bi araba* 'quite a big car', *kötü oluyo o bayağı* 'that is quite bad', *işte o fasiküller bayağı iyi öğretiyo* 'those booklets teach the topics quite good'.

Though the number of occurrences are limited, the results corroborate the previous observations that intensifiers are used to express stance in youth talk (Barbieri, 2008; Beltrama & Staum Casasanto, 2017; Núñez-Pertejo & Palacios-Martínez, 2018). Moreover, the findings point out that young speakers of Turkish display preferences for choosing between two semantically close intensifiers *aşırı* or *bayağı* in accordance with the characteristics of the affective domain in interaction.

This section has reviewed two types of intensifiers in the corpus: *amplifiers* and *downtoners*. The most frequently occurring intensifier was *çok* which corresponds to amplifier 'very' in English. In line with the arguments of the scholars who draw attention to the effect of speaker sex over the frequency of intensifier use (Fuchs, 2017; Lakoff, 1975; Murphy, 2010; Ito & Tagliamonte, 2003; Tagliamonte, 2005, 2008; Tagliamonte & Roberts, 2005), the analysis conducted in the CoTY also confirmed that the difference in terms of intensifier frequency was significant for the intensifier types (with a single exception of amplifier *fazla* 'excessively'). The results of the chi-square test showed that female speakers used intensifiers more than males with the exception of mother-plus swearing expressions, *gayet* 'excessively', *harbi(den)* 'really', and *öyle* 'so'. An additional analysis showed that the frequency of specific intensifiers vary based on the type of the speaker groups (female-female talk, male-male talk, mixed groups) as well. Later in this section, the attention was on the delexicalization process observed for a particular downtoner *bir tık* 'a bit' by adopting a diachronic corpus approach. The results suggest that over the course of a time period of about ten years, the lexical item *tık* transformed into *bir tık* as an intensifier which is saliently observed among Turkish speaking youth talk. Finally to conlude this section, the pragmatic differences of the intensifiers *aşırı* and *bayağı*, which both semantically correspond to 'excessively' in English, were examined. The analysis indicates that both intensifiers are used to state opinion about other people yet they have contrasting semantic prosody as in highlighting negative traits (i.e., *aşırı*) and positive traits (i.e., *bayağı*) of a person.

In this chapter thus far, the structural overview of the corpus, dominant topical and lexical characteristics of the data, and the most salient features of the four groups of interactional markers -*response tokens, vocatives, vague expressions, intensifiers*- were presented. In the following chapter, a summary of main findings together with the implications for further research will be provided.

CHAPTER 5

CONCLUSION

5.0 Presentation

In this chapter, following the purposes of the study, the major findings are summarized. Then, the directions and implications for future corpus studies focusing youth language will be presented.

5.1 Summary of Findings

This study had two complementary purposes. The first purpose was *to build* a sustainable tool to examine the linguistic practices of younger speakers of Turkish, and the second purpose was *to employ* this tool to explore the salient features of the spoken interaction between these speakers. In line with these purposes, the findings will be summarized under two layers: the corpus construction and the linguistic architecture.

5.1.1 Layer One: Corpus Construction

In terms of the first purpose of the study, a specialized spoken corpus, the Corpus of Turkish Youth Language (CoTY), was built. Covering the period of October 2019 to October 2021, a maximally representative sample was compiled by combining convenience sampling with maximal variation sampling. In line with the *participatory turn* in sociolinguistics and the action agenda proposed by the *open science movement*, this study made use of the contributory public participation model (Shirk, et al., 2012) to integrate an emic perspective into the data as well as to increase the data precision and accuracy.

The COVID-19 pandemic was a milestone which necessitated the data collection and sampling procedure to adapt to the new norms established for the social interactions and educational practices in the country. The national education was carried out by

distance education methods during the outbreak and many families in Turkey temporarily changed their cities of residence for facilitating the ease of living under pandemic circumstances. This situation shaped the scope of sampling frame of the CoTY and the modes of data the corpus covers. In terms of its sampling frame, an embracing approach was adopted to reach out to a wide range of residential locations in the country rather than focusing on a single province. As a result, the CoTY has a wider geographical coverage as it has obtained the spoken data of young speakers from 25 provinces across 12 regions in Turkey.

Given the current ease of access to global linguistic, semiotic, and cultural capital as well as the availability of various digital tools to maintain communication, it is not feasible to treat face-to-face and online modes of interaction as two separate interactional spheres. Rather, linguistic practices performed online are deeply embedded in the offline practices and norms of the communities (e.g., Androutsopoulos, 2006, 2008; Dovchin et al., 2018; Georgakopoulou, 2006, 2016; Page, 2018). This study advocates that this results in the emergence of an intricately interwoven nexus for the online and face-to-face modes of communication. As a result, the mode of communication was not a parameter with regard to the investigation of the interactional dynamics in the CoTY, although it was kept as metadata.

In addition to the mode of interaction, this study obtained a comprehensive account of metadata concerning the profile of the speakers (e.g., sex, socioeconomic status, school type) and the characteristics of communication (e.g., the frequency of communication, the setting, the ongoing activities during talk). Nevertheless, the foci of the analyses conducted in this study were not on revealing the influence of these categories over the identified patterns of linguistic practices, but to enable the researcher to situate the data in its authentic context as much as possible.

As the very first corpus compiled and constructed for Turkish youth talk, the CoTY comprises 168,748 tokens of 24,736 word types. The corpus data has 26 hours and 11 minutes of data which is naturally occurring and spontaneous speech collected in face-to-face or online informal contexts. The speaker groups consist of either 2 or 3 participants, and the speakers define their relationship with their interlocutor(s) to be 'friends'. In total, there are 123 unique speakers (62 females and 61 males) in the corpus

and the ages of the speakers range from 14 to 18, with 16-year-olds providing 38.5% of all data (64,927 tokens) in the whole corpus[43].

Though the corpus has a balanced distribution in terms of sex of the speakers, the distribution of the data in the corpus is inherently skewed to some extent. While female speakers provided 97,676 tokens, male speakers produced 71,072 tokens in the CoTY (corresponds to 58% and 42% of the whole data in the corpus, respectively). With regard to the distribution of data and the types of speaker groups, all-female interactions make up of 84,076 tokens (49.8% of the corpus), all-male interactions make up of 43,849 tokens (26%), and interaction which included both female and male speakers had 40,823 tokens (24.2%).

At first glance, although the CoTY seems to be smaller in terms of its current size compared to a number of available spoken youth talk corpora such as COLT, COLAm, and KiDKo, it stands out as a meticulously designed specialized corpus in terms of its clearly defined register characteristics and comprehensive metadata. The COLT, for instance, was reported to also include adult speakers and monologues (Stenström, 2002) as well as inconsistent and incomplete assignments of metadata (Stenström, 2013) in its structure. The CoTY, on the other hand, is made up of unscripted interactional data obtained exclusively from peers and rich metadata covering the profile of the speakers and their interactions. To provide a comparative view for the scope and the profile of the available youth talk corpora, Table 36 below presents their structural properties.

[43] This study recruited participants based on their high school grade levels and the balance between the number of speakers in each grade was set as the sampling criterion. In 2012, Primary Education Law no 6287 which is also known as '4+4+4 System' was implemented in Turkey. This law revised the starting age for primary education which resulted in the skewed distribution of ages across grades over the time. As a result of the aforementioned law, 16-year-old participants in the corpus has a wide spread across 9th, 10th and 11th grades.

Table 36 Profiles of spoken corpora for youth language

PROPERTIES	*COLT*	*COLA*	*Ph@ttSessionz*	*KiDKo*	*CORMA*	*JuBe*	*CoTY*
Full name	The Bergen Corpus of London Teenage Language	Corpus Oral de Lenguaje Adolescente	The Ph@ttSessionz speech database	Das Kiezdeutschkorpus	El Corpus Oral de Madrid	Jugendsprache Schweiz Korpus	The Corpus of Turkish Youth Language
Language	English	Spanish	German	mostly German	Spanish	German (Swiss)	Turkish
Age range	13-17	13-19	12-20	14-17	12-25	12-22	14-18
Speaker profile	adults + youth	adults + youth	youth	youth	youth	youth	youth
No. of speakers	31	145	864	23	139	26	123
Data collection	1993	2003-2007	2005-2007	2008-2015	2016-2019	2019-2021	2019-2021
Scope of metadata	limited & incomplete	limited: age, sex, grade, SES	limited: age, sex, dialect, COR	limited: age, sex, language(s)	comprehensive speaker metadata	limited: age, sex, education, ethnicity	comprehensive speaker and communication metadata
Hours	55	70	69	66	14	20	26
Tokens	444,166	463,047	110,000+ (utterances)	333,000	154,117	198,474	168,748
Register characteristics	spontaneous conversations + monologues	informal conversations	both read and non-scripted speech	spontaneous informal speech	spontaneous conversations	informal speech, some under observation	spontaneous informal conversations

5.1.2 Layer Two: Linguistic Architecture

In order to identify and explore the macro and micro characteristics of the linguistic architecture of Turkish youth talk, this study focused on *topical* and *lexical characteristics* of the interaction among the speakers of the corpus as well as four groups of *interactional markers.*

Topical characteristics were presented through coding all the topics and sub-topics mentioned within the conversations. A total of 47 conversational topics clustered under 11 main topics were identified in the CoTY. The identification of the types and the range of topics facilitated the contextualization of the corpus data, enabled the researcher to track the discursive strategies and lexical resources across different topics, and having an overview of the shared conceptual space of the speakers. The results showed that speakers conversed about a range of topics which covered a wide spectrum including daily topics such as shows they watch, schoolwork, daily chores as well as intimate and sensitive topics such as romantic relationships, politics, family problems, mental health, and issues of sexuality. The most frequently talked topics cluster around the main topic of *entertainment* (23%) which provides a shared conceptual space and a repertoire of linguistic and semiotic resources for speakers when jointly construct the interaction. The second most frequently mentioned topic is *social and emotional bonds* (20%) which led young speakers to demonstrate emotional engagement with their interlocutor by making use of various interactional markers. The third biggest cluster of topics was *education* (17%) which mainly consisted of the problems the speakers face in the education system, their study routines, and their academic goals and dreams. The saliency of this particular topic is the direct result of the profile of the participants as the sampling frame did not include young people who are NEETs or are in the active labour market but only recruited participants from young people actively enrolled in the education system in Turkey.

The lexical characteristics of the corpus were specified through generating the wordlist for the corpus and comparing the frequencies of these tokens with their frequencies in the reference corpus STC. Through utilizing this keyness analysis, positive *keywords* typical for Turkish youth talk were identified. The results yielded two groups of keywords: the first group consisted of *key concepts* which refer to nominals within the conceptual domains of *daily life* and *education*. The tokens in these domains are in line

with the distinctive topics in the corpus. The second group of keywords were the functions words which are labelled as ***interactional markers*** (after Ruhi, 2013) within the scope of this study. The linguistic entities in this group of keywords displayed socio-pragmatic functions in discourse and they were categorized into four: *response tokens, vocatives, vague expressions*, and *intensifiers*. In this study, each category of these interactional markers were presented and discussed within individual sub-sections. Though existing body of literature has highlighted various lexical characteristics of youth language such as pragmatic/discourse markers, intensifiers, reported speech, invariant tags, swear and taboo words among many others so far (see Androutsopoulos, 2010 for an overview), this study grounded its foci of investigation on the keyness analysis. In other words, the study adopted a *corpus-driven approach* to set the boundaries for the scope of distinctive features to be examined for the Turkish youth talk represented in the CoTY.

The first group of interactional markers were ***response tokens*** which demonstrate the active listenership behaviour of the interactants in groups. While the categorizations for English response tokens mainly made use of minimal and non-minimal distinction (Fellegy, 1995; Fishman, 1978; Gardner, 1997, 2001; Schegloff, 1982; Tottie, 1991), this study proposed a different categorization based on the morphological as well as pragmatic characteristics of Turkish. This categorization consisted of *non-lexical response tokens* which include short vocalizations and interjections, and *lexical response tokens* which included one-word lexical responses and small clusters of lexical responses such as repetitions of these responses and premodified responses.

All interactional markers in this study were identified using the EXAKT tool and KWIC analyses, and a total of 1305 ***non-lexical response tokens*** of 36 types and a total of 1728 ***lexical response tokens*** of 37 types were identified in the CoTY. The most frequently observed non-lexical response token was found to be *hı-hı* 'mm-hmm' (*AF*=337, *RF*=1997.06) and the most frequently observed lexical response token was *evet* 'yes' (*AF*=1582, *RF*=9374.93) in the corpus.

In addition to the conventional forms of response tokens, lexical items from the domains of taboo language (i.e., mother-plus swear words, vagina-related swearwords, and variants of 'fuck' in Turkish), originally religious expressions which may/may not have been used in that sense (i.e., *valla* 'really', *Allah* 'God', *inşallah* 'hopefully', *maşallah*

'wonderful', and *tövbe* as an expression of disbelief/disapproval') and contemporary slang (i.e., *harbi* 'really'*, aga be* 'come on bro'*, şaka* 'joke') were also utilized as response tokens by Turkish speaking youth. There were also forms of slang interjections such as *oha*, *yuh*, and *çüş* which correspond to English interjection *whoa* or in some cases *fuck-plus* variants are observed in the corpus. In order to investigate the pragmatic functions of response tokens in Turkish youth talk, the second most frequently occurring lexical response token *aynen '*exactly' (*AF*=329, *RF*= 1949,65) was selected as the foci of analysis.

A closer corpus analysis showed that the function of ***aynen* 'exactly'** goes beyond its traditionally prescribed adverbial use in Turkish youth talk. This lexical item prominently acts as a response token in the CoTY. The analysis based on the taxonomy of O'Keefe and Adolphs (2008) show that *aynen* is most frequently used as a ***continuer*** (47%) to maintain the flow of discourse and encourage the current speaker to continue to talk. In CoTY, *aynen* as a continuer token was also found to be a supportive marker of the co-construction of interactional humour among the speakers. The second most frequently function was ***convergence*** (25%) which was reported by the literature to mark the (dis)agreement and change of topics. The analysis of the instances of *aynen* as convergence tokens in the corpus, in addition, showed that this particular function is interwoven with facework, as well. The function of ***engagement*** (15%) operated on the affective level and enabled the listeners to display emotional engagement with the messages conveyed by their interlocutor. Finally, the smallest proportion of the tokens of *aynen* (13%) were identified as ***information receipt*** *tokens* which were used for confirming that the interlocutor understood an earlier account of content or message. Overall, the results proved that *aynen* has multiple pragmatic functions in spoken Turkish, and it is particularly salient in youth talk. This salience was identified though running corpus queries for *aynen* in other available Turkish corpora. The corpora used was the STC which provides spoken adult talk data in Turkish covering the period of 2008-2013 and the TNC which is a general corpus of contemporary written (and partially spoken) Turkish. Compared with the STC and the TNC data, *aynen* was more frequent in the CoTY (relative frequencies are *RF*=1949,65 in the CoTY, *RF*=195.53 in the STC, and *RF*=66.83 in the TNC), thus it was considered as a register-specific token for Turkish youth talk which can also be considered as a marker of a recent trend in spoken Turkish.

The second group of interactional markers examined in this study were ***vocatives***. The analysis revealed that speakers in the corpus made extensive use of this group of markers in terms of their overall token frequency and the number of types. Though studies previously underlined that vocatives are not used among close associates (Biber et al., 1999), recent work on youth language indicate that vocative use is a prominent characteristic of the interaction among young speakers even though their relationship is frequently identified as 'close associates' (Palacios Martínez, 2011, 2021; Parkinson, 2020; Rendle-Short, 2009, 2010; Roels et al., 2021; Stenström et al., 2002). The particular focus in this study was on *nominal vocatives* which excluded the forms of pronominal and personal names from its scope.

The corpus analysis yielded a total of 48 types of 2111 tokens of vocatives in the CoTY. The most frequently occurring vocative was *kanka* 'dude' and its variants *kanki, kank, kanks* (*AF*=680, *RF*=4029.67)*;* followed by *oğlum* (*AF*=452, *RF*=1789.65); and abi (*AF* =302, *RF*=1789.65). Among the observations made based on the corpus data, it was highlighted that categorizing the vocatives in terms of their original *semantic categories* (as proposed by Biber et al., 1999) such as 'endearment', 'familiarizer', and 'insult' offers a relatively narrow approach to explain the pragmatic functions of these tokens. Additionally, it was revealed that *insult vocatives* were used by both females and male speakers and these vocatives were observed in interactions among all types of speaker groups (female-female, male-male, and mixed) for pejorative as well as social bonding purposes. It is also noteworthy that this group of vocatives were the richest vocative category with regard to the number of types of tokens it had (*n*=14). Another observation was made with regard to the *referents* of the vocatives which showed that though a number of vocatives were semantically marked for gender in Turkish, the speakers used them to address both female and male addressees in the CoTY. All these observations suggested that vocatives in youth talk display *pragmatic extension* and thus, identifying their functions require a contextual and relational approach. For this purpose, the most frequently occurring vocative ***kanka* 'dude'** (*AF*=680, *RF*=4029.67) was examined by means of adopting McCarthy & O'Keeffe's (2003) taxonomy of *organizational* and *interpersonal* levels for the functions.

The analysis revealed that speakers use *kanka* relatively more for organizational purposes (*n*=374) compared to interpersonal purposes (*n*=306). As far as all the sub-functions are concerned, the analysis showed that kanka was used for all the sub-

functions in the corpora, namely *turn management, topic management, summons* under the organizational functions and *badinage, mitigator, relational* under the interpersonal functions. In terms of the sub-functions, vocative *kanka* was used for most frequently for the purpose of ***topic management*** (32%) which covers launching, expanding, shifting, closing the topic in interaction, followed by ***relational*** purpose (18%) which is used for conveying personal evaluations, agreements, face boosters, and ***mitigator*** purpose (15%) which attenuates the potential threats to directed at the positive/negative face of speakers in the corpus.

The pragmatic aspect of vocative use in youth talk was also discussed in relation to another phenomena observed in the corpus. It was found that 88% of the speakers use at least one type of vocative in their talk and 72% of speakers use more than one type of vocative (M=4.4, SD=3.04) for the same addressee in a single conversation. In this study, this particular phenomenon was discussed as ***address shifts.*** To examine this phenomenon further, the distribution of organizational and interpersonal functions of two semantically related vocatives *bro* and *kardeş* (both of which correspond to 'sibling' in English) was scrutinized. The analysis pointed a tendency of using *bro* for the function of topic management, and using *kardeş* for the function of badinage among the youth. The address shifts, then, exhibit the responsive nature of the vocatives with regard to the interactional goals young speakers aim to achieve. This section underlined that both the case of *kanka* and the phenomenon of address shifts support the argument that the pragmatic roles vocatives play in interaction go beyond the scope of their traditional categorizations such as familiarizer (i.e., *kanka*) or a fictive kinship vocative (i.e., *bro* and *kardeş*). In this vein, youth talk is found to be a rich data source to explore the responsive nature of vocatives in terms of their orientation towards the interactional goals a speaker wants to achieve.

The third group of interactional markers under investigation were ***vague expressions***. As markers utilized for projecting the shared conceptual space among the speakers, these expressions were previously noted for their presence in informal and intimate registers of talk (Clancy, 2016; Evison et al., 2007; Stenström et al., 2002). In this study, vague expressions were investigated under two groups: *vague references* and *vague additives*.

The analysis revealed 26 types of 4438 tokens of vague expressions in the corpus. ***Vague references*** constitute a bigger portion (68%) of the identified vague expressions. This group of vague expressions was divided into two sub-types. The first group of vague references was *indefinite references* denoting non-specific entities such as ***şey*** 'thing' which is the most frequently occurring vague expression in the whole corpus (*AF*=2093, *RF*=12403.11). In this first group, the referents of vague expressions was explored with a focus on *şey* 'thing'. The analysis demonstrated that the referent of *şey* 'thing' could be present in the same utterance, within the local co-text, the extended context, or the referent may not be present at all. In all those cases, the interaction was not disrupted due to the established 'shared knowledge' of the speakers. The second group was *generic references* which had *insan* 'one/person/human' (*AF*=21, *RF*=124.45) and *adam* 'man' (*AF*=5, *RF*=29.63) as the identified tokens. For this group of tokens, ***insan*** 'one' was discussed in terms of its functions in the domain of relational management. The results indicated that generic reference *insan* 'one' was generally used by speakers to convey a personal and often evaluative opinion towards the behaviours of an absent other.

As for ***vague additives***, the tokens were grouped into *approximators* and *general extenders.* While approximators were used to state a rough estimate of quantities or states such as *neredeyse* 'almost' (*AF*=22, *RF*=130.37), it was the general extenders which dominated this particular type of vagueness category (98% of the tokens were coded as general extenders). The analysis showed that adjunctive-disjunctive distinction was not valid for Turkish general extenders as they were flexible in terms of their pragmatic functions regardless of the *and-plus* or *or-plus* form they took. For Turkish, general extenders were found both in bare forms such as *vesaire* 'et cetera' (*AF*=3, *RF*=17.77) or as affixes such as reduplication marker *m-* (*AF*=16, *RF*=94.81). It was noted that even though it was spoken interaction, Turkish speaking youth use *vs* 'etc' which is the abbreviation of *vesaire* 'et cetera' used in written Turkish. By utilizing this ***truncated form***, young speakers appear to minimize the redundancy even further in their speech even though a vague expression is already in use for replacing an utterance or a series of utterances. Another observation noted was related to the ***m-reduplication*** process observed in youth talk. The analysis revealed that speakers apply a Turkish-specific morphological rule, *m*-reduplication to English words to generate general extenders. This innovative use of creating general extenders reflect the inherently dynamic nature of youth language which fully utilizes the available linguistic resources of the speakers.

In the final section of vague expressions, the study drew attention to the influence of the immediate context over the use of vague expressions. For this purpose, Biber et al. (2021) and Egbert et al.'s (2021) taxonomy of conversational communicative purposes of the discourse units in informal spoken interaction was utilized. The vague expression under investigation was determined as ***f(a)lan* 'and stuff'** (*AF*=1468, *RF*=8699.36) which was the most frequently occurring general extender in the CoTY. The analysis showed that *f(a)lan* was present in all conversational communicative purposes in the taxonomy, namely: (1) situation-dependent commentary, (2) joking around, (3) engaging in conflict, (4) figuring-things-out, (5) sharing feelings and evaluation, (6) giving advice and instructions, (7) describing or explaining the past, (8) describing or explaining the future, and (9) describing or explaining (time neutral). Confirming the existing literature that vague language is prominent in intimate and informal discourses (Channell, 1994; Clancy & McCarthy, 2015; Clancy, 2016; Cutting, 2001), general extender *f(a)lan* was identified more frequently in discourse units with the communicative purposes of *sharing feelings and evaluations* (corresponds to 31% of all the purposes). A number of distinct pragmatic functions of *f(a)lan* were observed in particular types of communicative purposes in the corpus. Among the most salient ones, the study highlighted the function of mitigation in *gossip talk* present in the episodes of ***sharing feelings and evaluations***, co-constructing *re-enactment* in the discourse units of ***describing or explaining the past***, and construction of *solidarity* in a future oriented hypothetical space within the discourse units of ***describing or explaining the future***. Adopting the aforementioned taxonomy offered a systematic approach to identify the relationship between the local context and functions of vague expressions.

The final group of interactional markers, ***intensifiers***, are the lexical items used for exaggerating or diminishing the message conveyed by the speakers. Their typical characteristics of productivity, expressiveness, recycling (Aijmer, 2018, 2020; Nevalainen & Rissanen, 2002; Stoffel, 1901; Tagliamonte 2008) are well suited with the innovative nature of youth talk. For the purposes of this study, the scope included adjectival and adverbial intensifiers along with taboo intensifiers. Following Biber et al. (1999), the binary categorization of *amplifiers* and *downtoners* were used to categorize intensifiers in the corpus.

Corpus queries yielded 29 types of 2856 tokens of intensifiers in the CoTY. ***Amplifiers*** which are used to intensify the strength of a particular aspect of the meaning of a lexical

item were found to be richer in terms of the types of intensifiers (*n*=24) and make up 93% of all intensifiers. Within this group, the most frequently occurring intensifier was *çok* 'very' (*AF*=1705, *RF*=10103.82) which is noted as a conventional amplifier in Turkish language. Following *çok* 'very', amplifier *bayağı* 'excessively' (*AF*=188, *RF*=1114.09) stands out as the register-specific keyword in the CoTY. In addition to *bayağı*, *aşırı* (*AF*=109, *RF*=645.93) is another intensifier and also a keyword which corresponds to 'excessively' in English. Though semantically similar, these intensifiers were found to display difference in terms of their ***semantic prosody***. While *aşırı* was used to underline the negative traits of a person, *bayağı* was used to highlight the positive traits.

Amplifiers in Turkish youth talk data included lexical items and phrases from the domain of ***taboo and swear words*** which were absent in the contemporary general spoken Turkish in the STC. These intensifiers were mother-plus swearing expressions *ana+* (*AF*=9, *RF*=53.33), vagina-plus swearing expressions *am+* (*AF*=2, *RF*=11.85), *manyak* 'crazy' (*AF*=5, *RF*=29.63), and *deli* 'lunatic' (*AF*=2, *RF*=11.85). Additionally, ***loan words*** such as *full* as a borrowing from English (*AF*=33, *RF*=195.56) and *süper* as an established anglicism of the word 'super' (*AF*=3, *RF*=17.78) were identified for their intensification functions. The functions of the amplifiers were highlighting emotion-laden messages, underlining personal opinions and expressing stance in youth talk.

Since a recurring research agenda for intensifiers to date has been the relationship between gender and the intensifier use, particular attention was given to the distribution of types and frequencies of intensifiers with regard to male and female speakers. It was found that the total number of intensifiers (*n*=1955) produced by female speakers is twice as much as that of male speakers (*n*=901) in the CoTY. Furthermore, with a single exception of amplifier *fazla* 'excessively', there was a significant difference between the frequencies of intensifiers used by females and males. Male speakers used mother-plus swearing expressions, *gayet* 'excessively', *harbi(den)* 'really', and *öyle* 'so' more frequently than females while female speakers used the rest of the intensifiers more frequently than male speakers. The type of speaker groups (i.e., female-female, male-male, mixed groups) also affected the frequency of intensifiers used. Except for *bayağı* 'excessively', *en* 'the most', *gayet* 'excessively', *full, vallahi* 'really', *harbiden* 'really', *öyle* 'so', and *manyak* 'crazy', all intensifiers were used more frequently in all-female groups.

In direct contrast with amplifiers, ***downtoners*** reduce the force of the message conveyed. As a result, hedging face threatening acts was the prominent function they display in the interaction among young speakers of Turkish. The most frequent downtoner in the CoTY is *biraz(cık)* 'barely' (*AF*=196, *RF*=1161.50) followed by ***bir tık* 'a bit'** (*AF*=26, *RF*=154.08). Particular attention was directed at the downtoner *bir tık* 'a bit' in order to trace an instance of delexicalization in Turkish. As a continuum, ***delexicalization*** refers to the process in which lexical items partly or fully lose their original meanings and transform into intensifiers (Partington, 1993; Tagliamonte & Roberts, 2005). A diachronic approach was adopted to explore the occurrences and the pragmatic uses of *bir tık* across Turkish youth talk, general spoken Turkish (i.e., the STC and spoken TNC), and general written Turkish (i.e., the TNC) corpora. Even though token *tık* was present in the STC (*AF*=12, *RF*=57.50), the cluster *bir tık* was not present in STC data at all. As an inanimate imitative form, bare *tık* was used to convey the meanings of 'quickly' or 'none'. As a complementary data source, the spoken part of TNC data showed that while *tık* (*AF*= 35, *RF*=34.52) exhibit the same meanings as in the STC, but provided no results in its spoken sub-corpus. The query in the written TNC yielded four occurrences of *bir tık* which were used as intensifiers just like in the CoTY. It is important to note that these instances belonged to data from blog posts published in 2012 and 2013. The analysis showed that the conversational style in the blogs reflected the informal spoken register of Turkish. As a consequence, the TNC data supported the argument that *tık* may have undergone delexicalization and shift into intensifier *bir tık* over the last decade.

As a contemporary slang intensifier in Turkish spoken by younger speakers, downtoner *bir tık* can be traced back to 2012 when the use of this emergent intensifier was reflected in language of young adults in the virtual space. Though limited in terms of the scope of data, this observation affirms the affordances of using corpus methods to explore the process of language change. Echoing Briz's (2003) words which suggest that "we are able to foresee the changes that our language may undergo in the future, since teenagers work like filters" (as cited in Jørgensen, 2013, p. 152), the CoTY and the future corpora studies lay solid ground for a more reusable, accountable, and transparent research which will be presented in the section that follows.

5.2 Future Directions

So far, the available spoken corpora of youth language have focused on the language of English (the COLT), Spanish (the COLAm and the CORMA), German (Ph@ttSessionz, KiDKo, and JuBe), and Nordic languages (the UNO and the ISLC). Reviewing these corpora and their research output, we see that adopting similar sampling frames (as in the COLT, the COLAm, and the UNO) enables cross linguistic comparability; having two corpora representing different periods of time of the same language allows diachronic comparison of linguistic practices in a speech community (as in the COLAm and the CORMA); and accessing corpora which focuses on speakers with different ethnic and social backgrounds can present accounts of translingual practices (as in KiDKo and JuBe). It is thanks to the affordances of corpus linguistics that the analyses are enhanced regarding the recent investigation foci of youth language. For Turkish, though, the CoTY represents the very first attempt to pave the way for any of the previously mentioned macro research agenda.

As there was neither an available corpus nor no predecessor corpus study on Turkish youth language, the study at hand aimed to lay a solid ground for the future studies within the intersection of youth language and corpus linguistics. Though youth language provides a wealth of data to explore, the difficulties of accessing participants below 18 years of age, obtaining naturally occurring and spontaneous speech data within the private domain of the speakers, and scarcity of methodologies used for systematic documentation and analysis of the data have resulted in the *absence of the youth* within the field of Turkish linguistics so far. Even though obtaining authentic language data from the young speakers and compiling a spoken corpus are labour-intensive, this study presents a roadmap through establishing the transparency of methodology and making data collection tools and procedures accessible to other researchers. Through prioritising and incrementing *open science practices*, it is hoped that emerging collaborative models for linguistics will add on to this body of work and will make youth language data more visible.

While exploring the topical and lexical characteristics of the CoTY, the analyses pinpointed several discursive observations that can inspire future scientific inquiries. Specific speech events, such as *conflict talk, gossip talk, conversational humour*, stood out as potential leads to follow to further scrutinize the interactional dynamics of jointly

constructed informal conversations among young speakers. Revealing the emergence of *identity work, facework, stylisation* within these speech events, for instance, will certainly contribute to the in-depth understanding of the multiple facets of interaction such as performativity, negotiation, and creativity observed in youth language. In this direction, the aforementioned speech events were annotated in the CoTY using Partitur-Editor tool of EXMARaLDA. Through this annotation, the goal is to generate sub-corpora of speech events or dominant discoursal characteristics in the CoTY. Furthermore, the interactional markers (i.e., *response tokens, vocatives, vague expressions, intensifiers*) investigated in the present study will be utilized as the baseline data to build the future discussions on the intersectional nature of youth interaction, as well.

As previously illustrated through a representative review of existing youth language literature, the dominance of research on English language prevails. Thus so far, several scholars have underlined the need for *cross-linguistic perspectives* on the discussions over the linguistic practices of the youth. In this sense, adopting a corpus approach is the inevitable and necessary answer to these calls. It is suggested that future youth language studies should diversify their scope in order to enable both intra- and cross-linguistic explorations. Though the literature mainly highlights the results of cross-linguistic comparisons concerning youth language, corpus linguistic studies also allow for *synchronic and diachronic analyses* of a target language. As was hinted by the diachronic exploration of the intensifier *bi tık* 'a bit' in this study, utilizing corpora of different registers or time periods of a language provides robust evidence for tracing linguistic variation and instances of innovation in a language. Additionally, it should be noted that developing youth corpora has vital implications on *first-language instruction*, as well as *foreign and second language pedagogy*. By incorporating them into language learning and teaching processes, language corpora of any type provide extensive opportunities for skills development, syllabus and material design.

At this point, it is highly vital to recall the *intertwined nature of online and offline linguistic practices*. The insights from the CoTY suggest that young speakers of Turkish manipulate the semiotic resources available in both their immediate physical context and the digital sphere while they are interacting. I argue that meaning making practices of the youth transcends the modes of communication employed. Future studies within this paradigm will certainly evolve our understanding of the pragmatics of language.

Rather than an end-product of a dissertation study, the CoTY is a long term project which is designed to grow in terms of size, the scope of sample, and modes of interaction over time. This dissertation study documents how a specialized corpus of Turkish youth language is compiled and what it offers for existing Turkish and cross-linguistic research. It is hoped that this project and complementary future studies by other researchers will contribute to the accumulation of methodological know-how in linguistics and build on the systematic and robust investigations into the linguistic practices of young people.

REFERENCES

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes Toward Open Science and Public Data Sharing: A Survey Among Members of the German Psychological Society. *Social Psychology, 50*(4), 252–60. https://doi.org/10.1027/1864-9335/a000384

Açıkalın, I. (1991). Türk gençlerinin yaşıtları arasında kullandıkları konuşma biçimleri. *Journal of Linguistic R*esearch, 2, 75–78.

Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal.* Routledge. https://doi.org/10.4324/9780203526149

Adolphs, S., & Knight, D. (2010). Building a spoken corpus. In A. A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 38-52). Routledge. https://doi.org/10.4324/9780203856949.ch4

Afful, J. B. A. (2006). Address terms among university students in Ghana: A case study. *Language and Intercultural Communication, 6(*1), 76-91. https://doi.org/10.1080/14708470608668909

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. John Benjamins. https://doi.org/10.1075/scl.10

Aijmer, K. (2011). Are you totally spy? A new intensifier in present-day American English. In S. Hancil (Ed.), *Marqueurs discursifs et subjectivité* (pp. 155–72). Universités de Rouen and Havre.

Aijmer, K. (2013). *Understanding pragmatic markers*. Edinburgh University Press. https://doi.org/10.1515/9780748635511

Aijmer, K. (2015). General extenders in learner language. In N. Groom, M. Charles, & S. John (Eds.), *Corpora, grammar and discourse* (pp. 211-33). John Benjamins. https://doi.org/10.1075/scl.73.10aij

Aijmer, K. (2018). That's well bad: Some new intensifiers in spoken in British English. In V. Brezina, R. Love, & K. Aijmer (Eds.), *Corpus approaches to contemporary British English* (pp. 60-95). Routledge. https://doi.org/10.4324/9781315268323-6

Aijmer, K. (2020). That's well good: A re-emergent intensifier in current British English. *Journal of English Linguistics, 49*(1), 18-38. https://doi.org/10.1177/0075424220979143

Aksan, M., & Demirhan, U. (2017). The verb bakmak 'to look' and its discourse functions: Corpus concordances. *Mersin University Journal of Linguistics and Literature, 14*(2), 85-107.

Aksan, Y., & Uçar, A. (2012). Türkçe Sözvarlığının Öğretiminde Derlem Kullanımı: Türkçe Ulusal Derlemi Örneği. In M. Aksan & Y. Aksan (Eds.), *Türkçe Öğretiminde Güncel Çalışmalar* (pp. 103-12). Mersin University Publishing.

Aksan, Y., Aksan, M., Mersinli, Ü., & Demirhan, U. U. (2017). *A frequency dictionary of Turkish*. Routledge. https://doi.org/10.4324/9781315733302

Aksoy, Ö. A. (1936). *Bir dili öğrenmek için en lüzümlü kelimeler ve bu kelimelerin belirtme usulü*. Gaziantep Halkevi Dil, Edebiyat, Tarih Şubesi Yayınları.

Akufuk, Z. (1996). *An Analysis of factors causing sutudent failure in foreign language learning at the Gaziantep Özel Sunguroğlu Erkek Lisesi and Cumhuriyet Lisesi* [Unpublished master's thesis]. Gaziantep University.

Alaca, S. (2014). *Televizyonda yayımlanan evlilik programlarındaki hitap şekilleri* [Unpublished master's thesis]. Hacettepe University.

Alemdaroğlu, A. (2007, August). *Formations of femininity at the intersection of class, gender and age: Young women in Turkey* [Paper presentation]. The Annual Meeting of American Sociological Association, New York, USA.

Alemdaroğlu, A. (2010). *Knowing your place: Inequalities, subjectivities and youth in Turkey* [Unpublished doctoral dissertation]. University of Cambridge.

Alkan Ataman, H. (2018). *Türkiye Türkçesinde hitaplar* [Unpublished doctoral dissertation]. Samsun Ondokuz Mayıs University.

Altunay, S., & Aksan, Y. (2018). Hayır and yok as pragmatic markers in Turkish: Findings from Spoken Turkish Corpus. *Mersin University Journal of Linguistics and Literature, 15*(2), 23-43.

Amador Moreno, C. P., McCarthy, M. J. & O'Keeffe, A. (2013). Can English provide a framework for Spanish response tokens? *The Yearbook of Pragmatics and Corpus Linguistics, 1*(1), 175-201. https://doi.org/10.1007/978-94-007-6250-3_9

Andersen, G. (1997). They like wanna see like how we talk and all that. The use of like as a discourse marker in London teenage speech. In M. Ljung (Ed.), *Corpus-based studies in English* (pp. 37-48). Rodopi.

Andersen, G. (1998). The pragmatic marker like from a relevance theoretical perspective. In A. Jucker & Y. Ziv (Eds.), *Discourse markers: Descriptions and theory* (pp. 147-70). John Benjamins. https://doi.org/10.1075/pbns.57.09and

Andersen, G. (2001). *Pragmatic markers and sociolinguistic variation.* John Benjamins. https://doi.org/10.1075/pbns.84

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics. 10*(4), 419-38. https://doi.org/10.1111/j.1467-9841.2006.00286.x

Androutsopoulos, J. (2007). Style online: Doing hip-hop on the German-speaking Web. In P. Auer (Ed.), *Style and social identities: Alternative approaches to linguistic heterogeneity* (pp. 279–317). Mouton de Gruyter.

Androutsopoulos, J. (2008). Potentials and limitations of discourse-centred online ethnography. *Language@Internet. 5*(8).

Androutsopoulos, J. (2015). Negotiating authenticities in mediatized times, *Discourse, Context & Media, 8,* 74–77. https://doi.org/10.1016/j.dcm.2015.06.003

Arslan, N., & Durukan, E. (2014). Yabancılara Türkçe öğretimi ders kitaplarında söz varlığı unsurlarının incelenmesi. *International Journal of Language Academy, 2*(4), 247-65. https://doi.org/10.18033/ijla.168

Aşık, U. (2007). *Yabancılar için temel Türkçe söz varlığı oluşturulması* [Unpublished master's thesis]. Dokuz Eylül University.

Athanasiadou, A. (2007). On the subjectivity of intensifiers. *Language Sciences, 29,* 554–65. https://doi.org/10.1016/j.langsci.2007.01.009

Aytaç-Demirçivi, K. (2021). *Backchannels in spoken Turkish.* [Unpublished doctoral dissertation]. Middle East Technical University.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum. https://doi.org/10.5040/9781350933996

Bal-Gezegin, B. (2013). How do we say no in Turkish? A corpus-based analysis of hayır and cık in Turkish. *Mersin University Journal of Linguistics and Literature, 10*(2), 53-73.

Balpınar, Z. (1996). The use of pronouns of power and solidarity in Turkish. *Journal of Linguistics Research*, 288-93.

Banguoğlu, T. (2011). *Türkçenin Grameri.* Türk Dil Kurumu Yayınları.

Barbieri, F. (2008). Patterns of age-based linguistic variation in American English. *Journal of Sociolinguistics, 12*(1), 58-88. https://doi.org/10.1111/j.1467-9841.2008.00353.x

Bauer, L., & Bauer, W. (2002). Adjective boosters in the English of young New Zealanders. *Journal of English Linguistics, 30*, 244–57. https://doi.org/10.1177/007524202030003002

Bayyurt, Y., & Bayraktaroglu, A. (2001). The use of pronouns and terms of address in Turkish service encounters. In A. Bayraktaroğlu & M. Sifianou (Eds.), *Linguistic Politeness across Boundaries: The case of Greek and Turkish* (pp.209-40). John Benjamins. https://doi.org/10.1075/pbns.88.09bay

Beltrama, A., & Staum Casasanto, L. (2017). Totally tall sounds totally younger: Intensification at the socio-semantics interface. *Journal of Sociolinguistics, 21*(2), 154-18. https://doi.org/10.1111/josl.12230

Berez-Kroeker, A. L., Gawne, L., Kelly, B. F., & Heston, T. (2017). A survey of current reproducibility practices in linguistics journals, 2003–2012. https://sites.google .com/a/hawaii.edu/data-citation/survey. Accessed March 17, 2020.

Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation, *Literary and Linguistic Computing*, *5*(4), 257–69. https://doi.org/10.1093/llc/5.4.257

Biber, D. (1993). Representativeness in corpus design, *Literary and Linguistic Computing, 8*(4), 243–57. https://doi.org/10.1093/llc/8.4.243

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263–86. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Egbert, J., Keller, D., & Wizner, S. (2021). Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. *Journal of Pragmatics, 171*, 20–35. https://doi.org/10.1016/j.pragma.2020.09.018

Biber, D., Johansson, S. Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* Longman.

Blanco-Suárez, Z. (2010, October). *On the origin and grammaticalisation of the intensifier deadly in English* [Paper presentation]. Langwidge Sandwidge Seminar, Manchester, UK.

Blum-Kulka, S. (2000). Gossipy events at family dinners: Negotiating sociability, presence and the moral order. In J. Coupland (Ed.), *Small talk* (pp. 213– 40). Pearson Education. https://doi.org/10.4324/9781315838328-12

Bodén, P. (2004). A new variety of Swedish? In S. Cassidy, F. Cox, R. Mannell, S. Palethorpe (Eds.), *Proceedings of the 10th Australian International Conference on Speech Science and Technology* (pp. 475–480), Macquarie University, Sydney, Australia.

Bodén, P. (2011). Adolescents' pronunciation in multilingual Malmö, Gothenburg and Stockholm. In R. Källström & I. Lindberg (Eds.), *Young urban Swedish: Variation and change in multilingual settings*, University of Gothenburg.

Bolinger, D. (1972). *Degree words.* Mouton. https://doi.org/10.1515/9783110877786

Bozkurt, B. (2015). *Türkçenin yabancı dil olarak öğretiminde orta seviye (B1-B2 Düzeyi) hedef sözcük belirleme çalışması* [Unpublished master's thesis]. Marmara University.

Braun, F. (1988). *Terms of address: Problems of patterns and usage in various languages and cultures*. Mouton De Gruyter. https://doi.org/10.1515/9783110848113

Brezina, V. (2018). *Statistics in Corpus Linguistics.* Cambridge University Press. https://doi.org/10.1017/9781316410899

Brown, G., & Yule, G. (1983). *Discourse Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511805226

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Brown, R., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style in Language* (pp. 253-276). MIT Press.

Brown, R., & Ford, M. (1961). Address in American English. *Journal of Abnormal and Social Psychology, 62*, 375-85. https://doi.org/10.1037/h0042862

Bulundu, A. C. (2022). *Level-based primary vocabulary in teaching Turkish as foreign language - A corpus based application* [Unpublished doctoral dissertation]. İnönü University.

Bulutay, T. (1969). *Türkiye'deki son öğrenci hareketleri üzerine bazı düşünceler*. Ankara Sosyal Bilimler Fakültesi Yayınları.

Burcu, E. (1997). *Çırak ve kalfa gençlik alt kültür grubu hakkında sosyolojik bir araştırma* [Unpublished doctoral dissertation]. Hacettepe University.

Burger S., & Schiel, F. (1998, May). RVG 1 – A database for regional variants of contemporary German. *Proceedings of LREC* (pp. 1083–1087). Granada, Spain.

Büyükkantarcığlu, N. (1992). İstanbul'daki üç ilkokulda farklı sosyal katmanlardan gelen üçüncü ve beşinci sınıf öğrencilerinde ölçüdil sözcük dağarcığı farklılıkları. *Gazi Eğitim Fakültesi Dergisi, 8*(3), 227-37.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Bybee, J. (2003). Mechanisms of change in grammaticization: The role of frequency. In B. Joseph & R. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 602–623). Blackwell. https://doi.org/10.1002/9780470756393.ch19

Canbulat, M. (2017). Türkçe öğretmen adaylarının gençlik diline ilişkin görüşlerinin toplumdilbilimsel yaklaşımla incelenmesi. *International Journal of Language Academy, 5*(14), 161–173. https://doi.org/10.18033/ijla.3506

Canbulat, M., Güven, Ö., & Çağlar, M. Ç. (2017). Youth language variation according to the levels of education. *International Journal of Language Academy*, *5*(September), 31–38. http://dx.doi.org/10.18033/ijla.3806

Çankaya, Ö. (1997). *The relationship among test anxiety, self-esteem and academic achievement in eleventh grade students* [Unpublished master's thesis]. Middle East Technical University.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–58. https://doi.org/10.1093/applin/16.2.141

Carter, R., & McCarthy, M. (1997). *Exploring Spoken English*. Cambridge University Press.

Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge University Press.

Çelik, K., & Lüküslü, D. (2010). Spotlighting a silent category of young females: The life experiences of 'House Girls' in Turkey, *Youth and Society, 44*(1), 28-48. https://doi.org/10.1177/0044118X10391636

Çelik, M. (1999). Da'nın işlevleri, *Dilbilim Araştırmaları, 10*, 25-32.

Çetintaş Yıldırım, F. (2018). Child-directed speech of Turkish mothers and fathers: A comparison based on the dimorphous verbal expression use. *Journal of Language Education and Research, 4*(1), 12-23.

Channell, J. (1994). *Vague language*. Oxford University Press.

Cheng, W. (2007). The use of vague language across spoken genres in an intercultural Hong Kong Corpus. In J. Cutting (Ed.), *Vague language explored* (pp. 161–81). Palgrave Macmillan. https://doi.org/10.1057/9780230627420_9

Cheng, W., & O'Keeffe , A. (2015). Vagueness. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 360-80). Cambridge University Press. https://doi.org/10.1017/CBO9781139057493.019

Cheshire, J. (2007). Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics, 11*(2), 155-93. https://doi.org/10.1111/j.1467-9841.2007.00317.x

Cheshire, J., & Williams, A. (2002). Information structure in male and female talk. *Journal of English Linguistics, 30(*2), 217–38. https://doi.org/10.1177/007242030002008

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics, 15*(2), 151-96. https://doi.org/10.1111/j.1467-9841.2011.00478.x

Chesire, J. (2013). Grammaticalisation in social context: The emergence of a new English pronoun. *Journal of Sociolinguistics, 17*(5), 608–33. https://doi.org/10.1111/josl.12053

Chesire, J., Nortier, J., & Adger, D. (2015) Emerging multiethnolects in Europe. *Occasional Papers Advancing Linguistics, 33,* 1–27.

Christa S., Brittan, D., Hodel, T., Hess, D., & Linder, A. (2021). *JuBe - Jugendsprache Schweiz Korpus (1.0)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5648157

Çiftçi, M. (1991). *Bir grup yükseköğretim öğrencisi üzerinde kelime serveti araştırması* [Unpublished master's thesis]. Afyon Kocatepe University.

Çıplak, M. (2005). *Uşak Merkez İlköğretim 5., 8. ve 11. sınıfların yazılı kelime hazinesinin belirlenmesi* [Unpublished master's thesis]. Afyon Kocatepe University.

Clancy, B. (2015). Hurry up baby son all the boys is finished their breakfast: Examining the use of vocatives as pragmatic markers in Irish Traveller and settled family discourse. In C. Amador Moreno, K. McCafferty, & E. Vaughan (Eds.), *Pragmatic markers in Irish English* (pp.228-47). John Benjamins. https://doi.org/10.1075/pbns.258.10cla

Clancy, B., & McCarthy, M. (2015). Co-constructed turn taking. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 430-53). Cambridge University Press. https://doi.org/10.1017/CBO9781139057493.023

Clancy, B., & O'Keeffe, A. (2015). Pragmatics. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of Corpus Linguistics* (pp. 231-51). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.014

Clancy, P. M., Thompson, S. A., Suzuki, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics, 26*(3), 355-87. https://doi.org/10.1016/0378-2166(95)00036-4

Clayman, S. E. (2012). Address terms in the organization of turns at talk: The case of pivotal turn extensions. *Journal of Pragmatics, 44*, 1853-67. https://doi.org/10.1016/j.pragma.2012.08.001

Coates, J. (2007). Talk in a play frame: More on laughter and intimacy. *Journal of Pragmatics, 39*, 29–49. https://doi.org/10.1016/j.pragma.2006.05.003

Coupland, J. (2000). Introduction: Sociolinguistic perspectives on small talk. In J.Coupland (Ed.), *Small talk* (pp. 1–25). Pearson Education. https://doi.org/10.4324/9781315838328-1

Coupland, J., & Jaworski, A. (2003). Transgression and intimacy in recreational talk narratives. *Research on Language and Social Interaction, 36,* 85–106. https://doi.org/10.1207/S15327973RLSI3601_5

Çöltekin, Ç., Doğruöz, A. S., & Çetinoğlu, Ö. (2022). Resources for Turkish natural language processing. *Lang Resources and Evaluation, 26,* 1-40. https://doi.org/10.1007/s10579-022-09605-4

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124-30. https://doi.org/10.1207/s15430421tip3903_2

Cutting, J. (2001). Speech acts of the in-group. *Journal of Pragmatics, 33*(8), 1207-33. https://doi.org/10.1016/S0378-2166(00)00056-4

Cutting, J. (2007). *Vague language explored*. Palgrave MacMillan. https://doi.org/10.1057/9780230627420

Cutting, J. (2011). Spoken discourse, In K. Hyland & B. Paltridge (Eds.), *Bloomsbury companion to discourse analysis* (pp.155-170). Bloomsbury.

Davaslıgil, Ü. (1980). Farklı sosyo-ekonmik ve kültürel çevreden gelen birinci sınıf çocukların dil gelişimine okulun etkisi. *İstanbul University Pedagogy Institute Journal of Pedagogy, 1*, 167-86.

Demez, G. (2009). Ergen kimliğinde düşler ve kabuslar: Taksim Beyoğlu civarında yaşayan gençlikte kimlik ve melezlik, In G. Putar (Ed.), *Kimlikler lütfen: Türkiye Cumhuriyeti'nde kimlik arayışı ve temsiliyeti* (pp. 323-36). METU Publishing.

Demir, A. (2017). *Kalkınma ajansı mali destek programlarının istatistiksel değerlendirilmesi: Ankara ili örneği* [Unpublished doctoral dissertation]. Ankara University.

Demir, G. (2012). The Development and current state of youth research in Turkey: An overview, *Young, 20(*1), 89-114. https://doi.org/10.1177/110330881102000105

Demir, G. (2021). *Gençlik algısı: 'Genç yaşam dünyası' ile ilgili gençler üzerine bir araştırma*. [Unpublished doctoral dissertation]. Marmara University.

Demir, N. (2010). Türkçede varyasyon üzerine. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Türkoloji Dergisi*, *2*, 93–106. https://doi.org/10.1501/Trkol_0000000191

Dempster, P., & Woods, D. (2011). The economic crisis through the eyes of Transana. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 12*(1). https://doi.org/10.17169/fqs-12.1.1515

Dickie, C., Schaeffler, F., Draxler, C., & Jinsch, K. (2009.) Speech recordings via the internet: An overview of the VOYS project in Scotland, *Proceedings of Interspeech, the 10th Annual Conference Of The International Speech Communication* (pp. 1807-1810), Brighton. https://doi.org/10.21437/Interspeech.2009-527

Dinçer, B., & Özaslan, M. (2004). *İlçelerin sosyo-ekonomik gelişmişlik sıralaması araştırması*. Kalkınma Bakanlığı, Bölgesel Gelişme ve Yapısal Uyum Genel Müdürlüğü, Ankara.

Dines, E. (1980). Variation in discourse – 'And stuff like that', *Language in Society, 9*, 13-31. https://doi.org/10.1017/S0047404500007764

Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language 1,* 19–80. https://doi.org/10.1075/sl.1.1.04dix

Doğan, İ. (1994). *Bir alt kültür olarak Ankara Yüksel Caddesi gençliği*. Ankara Kültür Bakanlığı Yayını.

Dorleijn, M., Mous, M., & Nortier, J. (2015). Urban youth styles in Kenya and the Netherlands. In J. Nortier & B. A. Svendsen (Eds.), *Language, youth and identity in the 21st Century: Linguistic practices across urban spaces* (pp. 271-89), Cambridge University Press. https://doi.org/10.1017/CBO9781139061896.019

Dovchin, S., Pennycook, A., & Sultana, S. (2018). *Popular culture, voice and linguistic diversity: Young adults on- and offline*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-61955-2

Drange, E.-M. (2002). Teenage slang in Norway. In L. E. Breivik & A. Hasselgren (Eds.), *From the COLT's mouth …and others': Language corpora studies in honour of Anna-Brita Stenström* (pp. 71-80). Brill. https://doi.org/10.1163/9789004334267_005

Drange, E.-M. (2009). Anglicisms in the informal speech of Norwegian and Chilean adolescents. In Stenström, A.-B. & A. Myre Jørgensen (Eds.), *Youngspeak in a multilingual perspective* (pp. 161-75). John Benjamins. https://doi.org/10.1075/pbns.184.12dra

Drange, E.-M., Hasund, I. K., & Stenström, A.-B. (2014). "Your mum!" Teenagers' swearing by mother in English, Spanish and Norwegian. *International Journal of Corpus Linguistics, 19*(1), 29–59. https://doi.org/10.1075/ijcl.19.1.02dra

Drummond, R. (2020). Teenage swearing in the UK. *English Worldwide, 41*(1), 59-88. https://doi.org/10.1075/eww.00040.dru

Drummond, R. (2018). Maybe it's a grime [t]ing. TH-stopping in urban British youth. *Language in Society, 47*(2), 171-96. https://doi.org/10.1017/S0047404517000999

Drummond, R. (2016). (Mis)interpreting urban youth language: White kids sounding black? *Journal of Youth Studies, 20*(5), 640-60. https://doi.org/10.1080/13676261.2016.1260692

Duncan, S., & Fiske, D.W. (1977). *Face-to-Face interaction: Research, methods, and theory* (1st ed.). Routledge. https://doi.org/10.4324/9781315660998

Duncan, S.D. Jr., & Niederehe, G. (1974). On signalling that it's your turn to speak. *Journal of Experimental Social Psychology, 10*(3), 234–47. https://doi.org/10.1016/0022-1031(74)90070-5

Dundes, A., Leach, J. W., & Özkok, B. (1972). The strategy of Turkish boys' verbal dueling rhymes. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 130–60). Holt, Rinewort and Winston.

Dynel, M. (2008). No aggression, only teasing. *Lodz Paper in Pragmatics, 4*, 241–61. https://doi.org/10.2478/v10016-008-0001-7

Dynel, M. (2009). Beyond a joke: Types of conversational humour. *Language and Linguistics Compass, 3(*5), 1284-99. https://doi.org/10.1111/j.1749-818X.2009.00152.x

Eckert, P. (1997). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 151–167). Oxford. https://doi.org/10.1002/9781405166256.ch9

Eckert, P. (2000). *Linguistic variation as a social practice.* Oxford.

Eder, D., & Enke, J. L. (1991). The structure of gossip: Opportunities and constraints on collective expression among adolescents. *American Sociological Review, 56,* 494–508. https://doi.org/10.2307/2096270

Efeoğlu, E. (2019, May). *Türkçe'de uyumun inşası: Sözlü iletişimde edimsel bir çerçevede 'canım' sözcüğü ve işlevleri* [Paper presentation]. 33rd National Linguistics Conference, Mersin, Turkey.

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge University Press. https://doi.org/10.1017/9781316584880

Egbert, J., Wizner, S., Keller, D., Biber, D., McEnery, T., & Baker, P. (2021). Identifying and describing functional discourse units in the BNC Spoken 2014. *Text & Talk, 41*(5-6), 715-37. https://doi.org/10.1515/text-2020-0053

Ehrnrooth, L. (2015). *FUSE - The Finnish Upper Secondary School Corpus of Spoken English.*

Emiroğlu, S. (2015). Öğrencilerin hazırlıksız konuşmada kullandığı farklı kelime sayısının tespiti ve incelemesi. *Journal of History School, 8*(23), 19-40. https://doi.org/10.14225/Joh764

Enghels, R., De Latte, F., & Roels, L. (2020). El Corpus Oral de Madrid (CORMA): Materiales para el estudio (socio)lingüístico del español coloquial actual. *Zeitschrift Fur Katalanistik, 33*, 45–76.

Erdinç Y. (2001). *Üniversite gençliğinin sosyo-kültürel profili üzerine bir alan araştırması: Gazi Üniversitesi örneği.* Gazi Üniversitesi İletişim Fakültesi Basımevi.

Erdoğan Baydilli, C. (1987). *Cinsiyet grupları açısından lise son sınıf öğrencilerinin sosyo-ekonomik konumları ve mesleksel tercihleri (Şarkışla Lisesi örneği)* [Unpublished master's thesis]. Cumhuriyet University.

Erdoğan, Y. (2013). Interactional functions of şey in Turkish: Evindence from Spoken Turkish Corpus. *Journal of Linguistics and Literature, 10*(2), 33-52.

Everts, E. (2003). Identifying a particular family humor style: A sociolinguistic discourse analysis. *International Journal of Humor Research, 16*(4), 369-412. https://doi.org/10.1515/humr.2003.021

Evison, J., McCarthy, M., & O'Keeffe, A. (2007). Looking out for love and all the rest of it: Vague category markers as shared social space. In J. Cutting (Ed.), *Vague language explored* (pp. 138–57). Palgrave Macmillan. https://doi.org/10.1057/9780230627420_8

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 17–47). Springer Open. http://doi.org/10.1007/978-3-319-00026-8.

Fellegy, A. (1995). Patterns and functions of minimal response. *American Speech, 70*, 186-99. https://doi.org/10.2307/455815

Fernanádez-Mallat, V. (2020). Forms of address in interaction: Evidence from Chilean Spanish. *Journal of Pragmatics, 161*, 95-106. https://doi.org/10.1016/j.pragma.2020.03.006

Fishman, P. (1978). Interaction: The work women do. *Social Problems, 25*, 397-406. https://doi.org/10.2307/800492

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional settings. In U. Connor & T. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11-33). John Benjamins. https://doi.org/10.1075/scl.16.02flo

Flowerdew, L. (2008). Corpora and context in professional writing. In V. Bhatia, J. Flowerdew & R.H. Jones (Eds.), *Advances in discourse studies* (pp. 115–127). Routledge.

Formentelli, M. (2007). The vocative mate in contemporary English: A corpus based study. In A.Sansò (Ed.), *Language resources and linguistic theory*, (pp. 180-99). Franco Angeli.

Fuchs, R. (2017). Do women (still) use more intensifiers than men? Compiling and analysing the Spoken British National Corpus 2014. *International Journal of Corpus Linguistics, 22*(3), 345–74. https://doi.org/10.1075/ijcl.22.3.03fuc

Furman, R., & Özyürek, A. (2007). Development of interactional discourse markers: Insights from Turkish children's and adults' oral narratives. *Journal of Pragmatics, 39,* 1742-57. https://doi.org/10.1016/j.pragma.2007.01.008

Gabrielatos, C. (2021). *Bibliography of discourse-oriented corpus studies*. Retrieved December 1, 2021, from http://ehu.ac.uk/docsbiblio.

Gardner, R. (1997). The conversation object mm - A weak and variable acknowledging token. *Research on Language and Social Interaction, 30*(2), 131-56. https://doi.org/10.1207/s15327973rlsi3002_2

Gardner, R. (1998). Between speaking and listening: the vocalisation of understandings. *Applied Linguistics, 19*(2), 204-24. https://doi.org/10.1093/applin/19.2.204

Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. John Benjamins. https://doi.org/10.1075/pbns.92

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice-Hall.

Gencan, T. N. (2007). *Dilbilgisi.* Ayraç Yayınevi.

Georgakopoulou-Nunes, A. (2016). Friendly comments: Interactional displays of alignment on Facebook and YouTube. In Leppänen, S., Kytölä, S. & Westinen, E. (Eds.), *Discourse and identification: Diversity and heterogeneity in social media practices* (pp. 178-207). Routledge.

Georgakopoulou, A. (2006). Postscript: Computer-mediated communication in sociolinguistics. *Journal of Sociolinguistics. 10*(4), 548–57. https://doi.org/10.1111/j.1467-9841.2006.00292.x

Georgakopoulou, A. (2007). *Small Stories, interaction and identities*. *Studies in narrative interaction 8*. John Benjamins. https://doi.org/10.1075/sin.8

Georgakopoulou, A. (2008). 'On MSN with buff boys': Self- and other-identity claims in the context of small stories1. *Journal of Sociolinguistics, 12*(5), 597–626. https://doi.org/10.1111/j.1467-9841.2008.00384.x

Georgakopoulou, A. (2016). From narrating the self to posting self(ies): A small stories approach to selfies. *Open Linguistics, 2*(1), 300–17. https://doi.org/10.1515/opli-2016-0014

Georgakopoulou, A. (2019). Designing stories on social media: A corpus-assisted critical perspective on the mismatches of story-curation. *Linguistics and Education, 62.* https://doi.org/10.1016/j.linged.2019.05.003

Georgakopoulou, A., & Charalambidou, A. (2011). Doing age and ageing- language, discourse and social interaction. In A. Aijmer & G. Andersen (Eds.), *Pragmatics of Society* (pp. 31-52). De Gruyter Mouton. https://doi.org/10.1515/9783110214420.31

Göçen, G., & Okur, A. (2016). Frequency and extensiveness of words in Turkish textbooks for foreigners. *Journal of National Education and Social Sciences, 45*(210), 447-77.

Goedertier, W., Goddijn, S., & Martens, J. (2000). Orthographic transcription of the Spoken Dutch Corpus. *Proceedings of the 2nd International Conference on Language Resources & Evaluation* (pp. 909-14). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2000/pdf/87.pdf

Goffman, E. (1967). On face-work: An analysis of ritual elements in social interaction. *Reflections, 4*(3),7-13. https://doi.org/10.1162/15241730360580159

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. Routledge. https://doi.org/10.4324/9780203340769

Gökşen, C. (2015). Anadolu'da erkeklerin eşleri için yaptıkları bazı adlandırmalar ve bunların sosyokültürel boyutları. *Türk Dili ve Edebiyatı Dergisi, 52*, 43-63.

Granger, S., Dupont, M., Meunier, F. Naets, H., & Paquot, M. (2020). *The International Corpus of Learner English. Version 3.* Presses Universitaires de Louvain.

Greckhamer, T., & Cilesiz, S. (2014). Rigor, transparency, evidence, and representation in discourse analysis: Challenges and recommendations. *International Journal of Qualitative Methods,* 422–43. https://doi.org/10.1177/160940691401300123

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, Volume 3, Speech acts* (pp. 41-58). Academic Press.

Gries, S., & Durrant, P. (2020). Analyzing co-occurrence data. In M. Paquot & S. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 141-59). Springer. https://doi.org/10.1007/978-3-030-46216-1_7

Guendouzi, J. (2001). You'll think we're always bitching: The functions of cooperativity and competition in women's gossip. *Discourse Studies, 3*, 29–51. https://doi.org/10.1177/1461445601003001002

Gumperz, J. (1982). *Discourse strategies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611834

Gunay, V. D. (2007). *Sözcükbilime Giriş.* Multilingual.

Günther, S. (2011). The dynamics of communicative practices in transmigrational contexts: 'insulting remarks' and 'stylized category animations' in everyday interactions among male youth in Germany. *Text & Talk, 31*(4), 447–73. https://doi.org/10.1515/text.2011.022

Hall, G. S. (1904). *Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, religion, and education* (Vols. I & II). D. Appleton & Co. https://doi.org/10.1037/10618-000

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 148-62). Longmans.

Halliday, M. A. K. (2003). On the 'architecture' of human language. In J. Webster (Ed.), *On language and linguistics, Volume 3, Collected works of M.A.K. Halliday* (pp. 15-16). Equinox.

Halupka-Rešetar, S., & Radić, B. (2003). Animal names used in addressing people in Serbian. *Journal of Pragmatics, 35*(12), 1891–1902. https://doi.org/10.1016/S0378-2166(03)00052-3

Harıt, Ö. (1971). *Kelime Hazinesi Araştırması.* Milli Eğitim Bakanlığı, Planlama Araştırma ve Koordinasyon Dairesi, Publication No: 103.

Hasselgren, A. (2000). The assessment of the English ability of young learners in Norwegian schools: an innovative approach. *Language Testing, 17*(2), 261–277. https://doi.org/10.1177/026553220001700209

Hasund, I. K., & Drange, E-M. (2014) English influence on Norwegian teenage slang. In J. Coleman (Ed.), *Global English slang: Methodologies and perspectives* (pp. 139-49). Routledge.

Hasund, I. K., & A.-B. Stenström. (1997). Conflict talk: A comparison of the verbal disputes of adolescent females in two corpora. In M. Ljung (Ed.), *Corpus-based studies in English* (pp. 119-33), Rodopi.

Hatipoğlu, Ç. (2008). Analysis of the social meanings of the second person pronoun SEN in Turkish. *Journal of Linguistics Research,* 15-29.

Hayran, T. (2019). *A study on the teaching of vocabulary in textbooks used in teaching foreign languages to foreigners* [Unpublished master's thesis]. Burdur Mehmet Akif Ersoy University.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 299-345). Cambridge University Press. https://doi.org/10.1017/CBO9780511665868.020

Hessner, T., & Gawlitzek, I. (2017). Totally or slightly different? Compiling and analysing the Spoken British National Corpus 2014. *International Journal of Corpus Linguistics, 22*(3), 403–28. https://doi.org/10.1075/ijcl.22.3.05hes

Heyd, T. (2014). Dude, Alter! A tale of two vocatives. *Pragmatics and Society, 5*(2), 271-95. https://doi.org/10.1075/ps.5.2.05hey

Housen, A. (2002). A corpus-based study of the L2-acquisition of the English verb system. Computer learner corpora, In S. Granger, J. H. Hung & S. Petch-Tyson (Eds.), *Second Language Acquisition and Foreign Language Teaching* (pp. 77- 116). John Benjamins. https://doi.org/10.1075/lllt.6.08hou

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524773

Hyland, K., & Paltridge, B. (2011). *Bloomsbury companion to discourse analysis.* Bloomsbury.

Ilbury, C. (2020). *Beyond the offline: Social media and the social meaning of variation in East London* [Unpublished doctoral dissertation]. Queen Mary University of London.

Ilbury, C. (2022a). Discourses of social media amongst youth: An ethnographic perspective. *Discourse, Context, and Media, 48*. https://doi.org/10.1016/j.dcm.2022.100625

Ilbury, C. (2022b). U Ok Hun?: The digital commodification of white woman style. *Journal of Sociolinguistics*. https://doi.org/10.1111/josl.12563

İMV-SAM, (1995). *Türkiye'de Gençlik, Türkiye'de 90'ların Gençliği*. İstanbul Mülkiyeliler Vakfı Sosyal Araştırmalar Merkezi, Yeni Yüzyıl Kitaplığı.

İpek Eğilmez, N. (2010). *İlköğretim Türkçe ders kitaplarındaki söz varlığının ilköğretim dördüncü sınıf öğrencilerinin yazılı anlatımlarına aktarımı* [Unpublished doctoral dissertation]. Bursa Uludağ University.

İpek, B. (2016). Türkiye Türkçesinde sözcük türleriyle yapılan pekiştirme. *Karadeniz Sosyal Bilimler Dergisi, 32*(32), 145-56. https://doi.org/10.17498/kdeniz.279536

İpekçi, A. (2005). *İlköğretim 7. Sınıf öğrencilerinin kelime serveti üzerine bir Araştırma* [Unpublished master's thesis]. Bolu Abant İzzet Baysal University.

Irwin, P. (2014). So [totally] speaker-oriented: An analysis of 'Drama SO'. In R. Zanuttini & L. R. Horn (Eds.), *Microsyntactic variation in North American English* (pp. 29-70). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199367221.003.0002

Işık-Güler, H., & Eröz-Tuğa, B. (2017). Sözlü Türkçe Derlemi ve Türkçe Ilusal Derlemi'nde (u)lan'in edimbilimsel bir incelemesi. *Mersin University Journal of Linguistics and Literature*, *14*(2), 37-60.

Ito, R., & Tagliamonte, S. (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers, *Language in Society, 32*(2), 257-79. https://doi.org/10.1017/S0047404503322055

Jasanoff, S. S. (2003). Technologies of humility: Citizen participation in governing science. *Minerva, 41*(3), 223–44. https://doi.org/10.1023/A:1025557512320

Jaworski, A., & Galasiński, D. (2000). Vocative address forms and ideological legitimisation in political debates. *Discourse Studies, 2*(1), 65-83. https://doi.org/10.1177/1461445600002001002

Jaworski, A., & Coupland, J. (2005). Othering in gossip: "You go out you have a laugh and you can pull yeah okay but like...", *Language in Society, 34*(5), 667–94. https://doi.org/10.1017/S0047404505050256

Jaworski, A., & Coupland, N. (2006). *The discourse reader* (2nd ed.). Routledge.

Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapping tag-positioned address terms in closing sequences. *Semiotica, 9*, 47-96. https://doi.org/10.1515/semi.1973.9.1.47

Jefferson, G. (1984). Notes on systematic deployment of the acknowledgment tokens 'yeah' and 'mm hm'. *Papers in Linguistics, 17*(2), 197-216. https://doi.org/10.1080/08351818409389201

Jefferson, G. (1990). List construction as a task and resource. In G. Psathas (Ed.), *Interaction Competence* (pp. 63-92). University Press of America.

Jonsson, R. (2018). Swedes can't swear: Making fun at a multiethnic secondary school. *Journal of Language, Identity & Education, 17*(5), 320-35. https://doi.org/10.1080/15348458.2018.1469412

Jørgensen, A. M. (2008). A matter of politeness? A contrastive study of phatic talk in teenage conversation. *Special Issue of Pragmatics, 18*(4), 635–657.

Jørgensen, A. M. (2009). Introduction. In A.-B. Stenström, & A. M. Jørgensen, (Eds.), *Youngspeak in a multilingual perspective*. Benjamin's Publishing Company.

Jørgensen, A. M. (2010). *Love Ya Hate Ya: The sociolinguistic study of youth language and youth identities*. Cambridge Scholars Publishing.

Jørgensen, A. M. (2013). Spanish teenage language and the COLAm-corpus. *Bergen Language and Linguistics Studies, 3*(1), 151-66. https://doi.org/10.15845/bells.v3i1.368

Jørgensen, J. N. (2008). Polylingual languaging around and among children and adolescents. *International Journal of Multilingualism, 5*(3), 161-76. https://doi.org/10.1080/14790710802387562

Kafes, H. (1998). *Enhancing EFL students reading comprehension through conciousness-raising training on contextual guesswork* [Unpublished master's thesis]. Anadolu University.

Kağıtçıbaşı, Ç. (1994). Psychology in Turkey. *International Journal of Psychology, 29*(6), 729–38. https://doi.org/10.1080/00207599408246562

Kağıtçıbaşı, Ç. (1996). The Autonomous-relational Self: A new synthesis, *European Psychologist 1*(3), 180–86. https://doi.org/10.1027/1016-9040.1.3.180

Kaplan, İ. (1999). *Türkiye'de milli eğitim ideolojisi*. İletişim Publishing.

Karaağaç, G. (2013). *Dil Bilimi Terimleri Sözlüğü.* TDK Yayınları.

Karadağ, Ö. (2005). *İlköğretim I. kademe öğrencilerinin kelime hazinesi üzerine bir araştırma* [Unpublished doctoral dissertation]. Gazi University.

Karaşin, H. (2008). Bağlaç görevinde kullanılan ya'nın anlamsal işlevler. *U.Ü. Fen-Edebiyat Fakültesi Sosyal Bilimler Dergisi, 9*(14), 215-227.

Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing Into Open Science: A Guide for Graduate Students and Their Advisors. *Psychology, 7*(1), 18684. https://doi.org/10.1525/collabra.18684

Kaynarpınar, F. (2021). *Approval markers in Turkish: A corpus-driven study* [Unpublished master's thesis]. Mersin University.

Kentel, F. (1995). *Türkiye'de 1990'ların Gençliği*. Yeni Yüzyıl Kitaplığı.

Kiesling, S. F. (2004). Dude. *American Speech, 79*(3), 281-305. https://doi.org/10.1215/00031283-79-3-281

Kilgariff, A., Rundell, M.. & Uí Dhonnchadha, E. (2006). Efficient corpus development for lexicography: Building the New Corpus for Ireland. *Language Resources and Evaluation, 40*, 127–52. https://doi.org/10.1007/s10579-006-9011-7

Kırık, A. M. (2012). Sosyal medyada gençlerin dili kullanımı ve yozlaşma problemi. In *Proceedings of the International Symposium on Language and Communication: Research Trends and Challenges* (pp. 1017-1030). İzmir University, Turkey.

Kışlalı, A. T. (1972). *Öğrenci ayaklanmalarının nedenleri ve rejim sorunu ile ilişkisi*. Hacettepe University Social Sciences Institute.

Kışlalı, A. T. (1974). *Öğrenci ayaklanmaları.* Bilgi Publishing.

Koçak, H. (1999). *Sağlık Meslek Lisesi öğrencilerinin kelime hazinesi ile ilgili bir araştırma* [Unpublished master's thesis]. Gazi University.

Koester, A. (2007). About twelve thousand or so: Vagueness in North American and UK offices. In J. Cutting (Ed.), *Vague language explored* (pp. 40-61). Palgrave Macmillan. https://doi.org/10.1057/9780230627420_3

Koester, A. (2010). Building small specialised corpora. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66-79). Routledge. https://doi.org/10.4324/9780203856949.ch6

König, G. (1990). Türkçe'de sen/siz adıllarının ikinci tekil şahıs için kullanımına toplumdilbilimsel bir yaklaşım. In A. S. Özsoy & H. Sebütekin (Eds.), *Proceedings of 4th Linguistics Symposium* (pp. 175-84). Boğaziçi University, Turkey.

Konrad Adenauer Foundation, (1999). *Türk Gençliği 98: Suskun kitle büyüteç altında*. Konrad Adenauer Foundation.

Köprülü, Ö. (1994). *Psycholinguistic factors on English learning of students 11 to 13 in Avni Akyol High School* [Unpublished master's thesis]. Dokuz Eylül University.

Korkmaz, Z. (2003). *Türkiye Türkçesinin Grameri.* TDK Yayınları.

Kotsinas, U.-B. (1988). Immigrant children's Swedish: A new variety? *Journal of Multilingual and Multicultural Development, 9,* 129-140. https://doi.org/10.1080/01434632.1988.9994324

Kozanoğlu, C. (1992). *Cilalı imaj devri: 1980'lerden 90'lara Türkiye ve starları*. İletişim Publishing.

Kozanoğlu, H. (1993). *Yuppieler, prensler ve bizim kuşak.* İletişim Publishing.

Küntay, E. (1999). İstanbul'da sokak çocukları: 1963-1998 arasında bir değerlendirme. In B. Onur (Ed.), *Republic and Children:* 2nd *National Conference on Children Culture* (pp. 570-81). Ankara University Children Culture Research and Application Centre Publishing.

Kurtoğlu, Ö., & Uçar, A. (2011). Aspects of vocabulary in Turkish language course book corpus. In V. D. Günay, Ö. Fidan, B. Çetin & F. Yıldız (Eds.), *Türkçe Öğretimi Üzerine Çalışmalar* (pp. 408-420). Dokuz Eylül University.

Kurudayıoğlu, M. (2005). *İlköğretim II. kademe öğrencilerinin kelime hazinesi üzerine bir Araştırma* [Unpublished doctoral dissertation]. Gazi University.

Labov, T. (1982). Social structure and peer terminology in a black adolescent gang. *Language and Society, 2*, 391–411. https://doi.org/10.1017/S0047404500009374

Labov, W. (1966). *The social stratification of English in New York City.* Center for Applied Linguistics.

Labov, W. (1985). Intensity. In D. Schiffrin (Ed.), *Meaning, form and use in context: Linguistic applications* (pp. 43-70), Georgetown University Press.

Labov, W. (2001). *Principles of linguistic change, Vol 2: Social factors.* Blackwell.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2*(4), 458-508. https://doi.org/10.1007/bf00262952

Lakoff, R. (1975). *Language and woman's place*. Harper & Row.

Lam, P. (2009). The making of a BNC customised spoken corpus for comparative purposes. *Corpora, 4*(1), 167–88. https://doi.org/10.3366/E174950320900029X

Lampert, M. D., & Ervin-Tripp, S. M. (2006). Risky laughter: Teasing and self-directed joking among male and female friends. *Journal of Pragmatics, 38*(1), 51–72. https://doi.org/10.1016/j.pragma.2005.06.004

Leech, G. (1999). The distribution and function of vocatives in American and British English conversation. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 107-18). Rodopi.

Leech, G. (2007). New resources, or just better old ones? In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 134–49). Rodopi.

Levey, S. (2012). General extenders and grammaticalization: Insights from London preadolescents. *Applied Linguistics, 33*(3), 257-81. https://doi.org/10.1093/applin/ams003

Levinson, S. (1983). *Pragmatics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511813313

Lewis, G.L. (2000). *Turkish Grammar.* Oxford University Press.

Lorenz, G. (2002). Really worthwhile or not really significant? A corpus-based approach to the delexicalization and grammaticalization of intensifiers in Modern English. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 143-61). John Benjamins. https://doi.org/10.1075/tsl.49.11lor

Love, R. (2017). *Bad language revisited: Swearing in the Spoken BNC2014.* Paper presented at Corpus Linguistics 2017 Conference, University of Birmingham.

Lukkari, L. (2020). *L1 use in EFL oral proficiency tests* [Unpublished master's thesis]. University of Helsinki.

Lüküslü, D. (2009). *Türkiye'de gençlik Miti: 1980 sonrası Türkiye gençliği.* İletişim Publishing.

Macaulay, R. (1995). The adverbs of authority. *English World-Wide, 16*, 37-60. https://doi.org/10.1075/eww.16.1.03mac

Macaulay, R. (2002). Extremely interesting. Very interesting, or only quite interesting? Adverbs and social class. *Journal of Sociolinguistics, 6*(3), 398-417. https://doi.org/10.1111/1467-9481.00194

Macaulay, R. (2006). Pure grammaticalization: The Development of a teenage intensifier. *Language Variation and Change, 18*, 267-83. https://doi.org/10.1017/S0954394506060133

Madsen, L. M. (2015). *Fighters, girls and other identities: Sociolinguistics in a martial arts club*, Multilingual Matters.

McCarthy, M. (2002). Good listenership made plain: British and American non-minimal response tokens in everyday conversation. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 49–71). John Benjamins. https://doi.org/10.1075/scl.9.05mcc

McCarthy, M. (2003). Talking back: 'Small' interactional response tokens in everyday conversation. *Research on Language and Social Interaction, 36*(1), 33–63. https://doi.org/10.1207/S15327973RLSI3601_3

McCarthy, M., & Carter, R. (2001). Size isn't everything: Spoken English, corpus, and the classroom, *TESOL Quarterly, 35*(2), 337–40. https://doi.org/10.2307/3587654

McCarthy, M., & O'Keeffe, A. (2003). 'What's in a Name?': Vocatives in casual conversations and radio phone-in calls. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 153–185). Rodopi. https://doi.org/10.1163/9789004334410_010

McCarthy, M., & O'Keeffe, A. (2002, September). *Vague language and participation framework: Indices of identity among group, culture and nation* [Paper presentation]. The 35th Annual Conference of the British Association for Applied Linguistics, University of Cardiff, UK.

McEnery, T. (2005/2012). *Corpus Linguistics. Oxford Handbooks Online*. Oxford. https://doi.org/10.1093/oxfordhb/9780199276349.013.0024

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics : Method, theory and practice*. Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

Méndez-Naya, B. (2003). On intensifiers and grammaticalization: The case of SWIÞE. *English Studies, 84(*4), 372-91. https://doi.org/10.1076/enst.84.4.372.17388

Méndez-Naya, B. (2008). Special issue on English intensifers. *English Language and Linguistics, 12,* 213-19. https://doi.org/10.1017/S1360674308002591

Miller, A. (2020). Analyzing frequency lists. In M. Paquot & S. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 77-97). Springer. https://doi.org/10.1007/978-3-030-46216-1_4

Murphy, B. (2010). *Corpus and Sociolinguistics. Investigating age and gender in female talk*. John Benjamins. https://doi.org/10.1075/scl.38

Napoli, M., & Ravetto, M. (2017). Ways to intensify: Types of intensified meanings in Italian and German. In M. Napoli & M. Ravetto (Eds.), *Exploring intensification: Synchronic, diachronic & cross-linguistic perspectives* (pp.327-52). John Benjamins. https://doi.org/10.1075/slcs.189.17nap

Nevalainen, T., & Rissanen, M. (2002). Fairly pretty or pretty fair? On the development and grammaticalization of English downtoners. *Language Sciences, 24*(3), 359-80. https://doi.org/10.1016/S0388-0001(01)00038-9

Neyzi, L. (2001). Object or subject? The paradox of 'Youth' in Turkey, *International Journal of Middle East Studies, 33*(13), 411-32. https://doi.org/10.1017/S002074380100304X

Nişanyan, S. (2012). *Sözlerin soyağacı: Çağdaş Türkçenin etimolojik sözlüğü.* Everest.

Norrby, C., & Winter, J. (2002). Affiliation in adolescents' use of discourse extenders. In C. Allen (Ed.), *Proceedings of the 2001 Conference of the Australian Linguistic Society*. http://www.als.asn.au/proceedings/als2001/winter_norrby.pdf

Norrick, N. R. (1993). *Conversational joking: humour in everyday talk*. Indiana University Press.

Norrick, N. R. (2009). A theory of humour in interaction. *Journal of Literary Theory, 3*(2), 261-84. https://doi.org/10.1515/JLT.2009.015

Nortier, J. (2016). Characterizing urban youth speech styles in Utrecht and on the Internet. *Journal of Language Contact, 9*(1), 163–85. https://doi.org/10.1163/19552629-00901007

Núñez Pertejo, P., & Palacios Martínez, I. M. (2014). That's absolutely crap, totally rubbish: The use of the intensifiers absolutely and totally in the spoken language of British adults and teenagers. *Functions of Language, 21*(2), 210–237. https://doi.org/10.1075/fol.21.2.03pal

Núñez-Pertejo, P., & Palacios-Martínez, I. (2018). Intensifiers in MLE: New trends and developments. *Nordic Journal of English Studies, 17*(2), 116-55. https://doi.org/10.35360/njes.436

O'Keeffe, A. (2004). 'Like the wise virgins and all that jazz': Using a corpus to examine vague language and shared knowledge. *Language and Computers, 52*(1), 1–20. https://doi.org/10.1163/9789004333772_002

O'Keeffe, A., & Adolphs, S. (2008). Response tokens in British and Irish discourse: Corpus, context and variational pragmatics. In K. P. Schneider & A. Barron (Eds.), *Variational Pragmatics* (pp. 69–98). John Benjamins. https://doi.org/10.1075/pbns.178.05ok

Obuz, F. G. (2012). *İlköğretim 8. sınıf öğrencilerinin sözlü anlatımlardaki aktif kelime serveti üzerine bir araştırma (Uşak/Ulubey Örneği)* [Unpublished master's thesis]. Uşak University.

Oflazer, K., Say B., Hakkani-Tür, D.Z, & Tür, G. (2003). Building a Turkish Treebank. In A. Abeillé (Ed.), *Treebanks: Building and using parsed* corpora (pp. 261-277). Kluwer Academic. https://doi.org/10.1007/978-94-010-0201-1_15

Oswalt, R. L. (1994). Inanimate imitatives in English. In L. Hinton, J. Nichols, & J. Ohala, (Eds.), *Sound Symbolism* (pp. 293-308). Cambridge University Press. https://doi.org/10.1017/CBO9780511751806.020

Overstreet, M. (1999). *Whales, candlelight and stuff like that: General extenders in English discourse*. Oxford University Press.

Overstreet, M. (2005). And stuff un so: Investigating pragmatic expressions in English and German. *Journal of Pragmatics, 37,* 1845-64. https://doi.org/10.1016/j.pragma.2005.02.015

Overstreet, M. (2012). Vagueness and hedging. In G. Andersen & K. Aijmer (Eds.), *Pragmatics and Society* (pp. 293-320). De Gruyter Mouton. https://doi.org/10.1515/9783110214420.293

Overstreet, M., & Yule, G. (2002). The metapragmatics of and everything. *Journal of Pragmatics, 34*. 785–94. https://doi.org/10.1016/S0378-2166(01)00036-4

Overstreet, M., & Yule, G. (1997). On being inexplicit and stuff in contemporary American English. *Journal of English Linguistics, 25*(3), 250–58. https://doi.org/10.1177/007542429702500307

Oyetade, S. O. (1995). A sociolinguistic analysis of address forms in Yoruba. *Language in Society, 24*, 515-535. https://doi.org/10.1017/S004740450001900X

Özaslan, M., Dinçer, B., & Özgür, H. (2006, September). *Regional disparities and territorial indicators in Turkey: Socio-Economic Development Index (SEDI)* [Paper presentation]. European Regional Science Association Conference, Volos, Greece.

Özbay, M., & İpek, O. (2015). Ortaokul öğrencilerinin kullandıkları arkadaşlık hitaplarının konuşma eğitimi açısından değerlendirilmesi. *Dil ve Edebiyat Eğitimi Dergisi, 15*, 90-105.

Özbek, N. (1995). *Discourse markers in Turkish and English: A comparative study* [Unpublished doctoral dissertation]. Nottingham University.

Özbek, N. (1998). Yani, işte, şey, ya: Interactional markers of Turkish. In A. Goksel & C. Kerslake (Eds.), *Proceedings of the 9th Interactional Conference on Turkish Linguistics*. Harrassowitz Publishing House.

Özcan, F. H. (2016). Choice of address terms in conversational setting. *International Journal of Human Sciences, 13*(1), 982-1002. https://doi.org/10.14687/ijhs.v13i1.3489

Özcan, G. (2015). *A corpus-driven analysis of evet 'yes' and hı-hı in Turkish: Evidence from the spoken Turkish corpus* [Unpublished master's thesis] Mersin University.

Özensel, M. (2009). *Türkiye'deki gençlik ve spor ile ilgili kamu kurumları ve sivil toplum kuruluşlarının gençlik politikalarının değerlendirilmesi ve bir model örneği* [Unpublished doctoral dissertation]. Abant İzzet Baysal University.

Özer, N. (2019). *Address terms in Turkish: A corpus-based study* [Unpublished master's thesis]. Mersin University.

Özgen, M., & Karataş, B. (2018). The distributional and categorical properties of the lexical item falan. *Dil Dergisi, 169*(2), 39-66.

Page, R. (2018). Group selfies and Snapchat: From sociality to synthetic collectivisation. *Discourse, Context & Media, 28(*1), 79-92. https://doi.org/10.1016/j.dcm.2018.10.003

Palacios Martínez, I. M. (2011a). The language of British teenagers. A preliminary study of its main grammatical features. *Atlantis, 33*(1), 105-126.

Palacios Martínez, I. M. (2011b). 'I might, I might go I mean it depends on money things and stuff' A preliminary analysis of general extenders in British teenagers' discourse. *Journal of Pragmatics, 43*, 2452-70. https://doi.org/10.1016/j.pragma.2011.02.011

Palacios Martínez, I. M. (2011c). The language of British Teenagers. A preliminary study of its main grammatical features. *Journal of the Spanish Association of Anglo-American Studies, 33*(1), 105-26.

Palacios Martínez, I. M. (2018). 'Help me move to that, blood'. A corpus-based study of the syntax and pragmatics of vocatives in the language of British teenagers. *Journal of Pragmatics, 130*, 33-50. https://doi.org/10.1016/j.pragma.2018.04.001

Palacios Martínez, I. M. (2021). Taboo vocatives in the language of London teenagers. *Pragmatics, 31*(2), 250-277. https://doi.org/10.1075/prag.19028.pal

Palacios Martínez, I. M., & Núñez Pertejo, P. (2012). He's absolutely massive. It's a super day. Madonna, she is a wicked singer. Youth language and intensification: A corpus-based study. *Text & Talk, 32*(6), 773–96. https://doi.org/10.1515/text-2012-0036

Palacios Martínez, I. M., & Núñez Pertejo, P. (2014). Strategies used by English and Spanish teenagers to intensify language: A contrastive corpus-based study. *Spanish in Context, 11*(2), 175–201. https://doi.org/10.1075/sic.11.2.02pal

Palacios Martínez, I. M., & Núñez Pertejo, P. (2015). 'Go up to miss thingy'. 'He's probably like a whatsit or something'. Placeholders in focus. The differences in use between teenagers and adults in spoken English. *Pragmatics, 25*(3), 425-51. https://doi.org/10.1075/prag.25.3.05pal

Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics, 12* (1), 47-65. https://doi.org/10.1515/cogl.12.1.47

Paradis, C. (2008). Configurations, construals and change: Expressions of degree. *English Language and Linguistics, 12*(2), 317–43. https://doi.org/10.1017/S1360674308002645

Parkinson, J. (2020). On the use of the address terms guys and mate in educational context. *Journal of Pragmatics, 161*, 81-94. https://doi.org/10.1016/j.pragma.2020.03.003

Pars, V. B., & Pars, C. B. (1954). *Okuma psikolojisi ve ilkokuma öğretimi.* Maarif Publication.

Partington, A. (1993). Corpus Evidence of language change: The case of intensifiers. In M. Baker, G. Francis, & A. Tognini-Bonelli (Eds.), *Text and Technology. In honour of John Sinclair* (pp. 177-92). John Benjamins. https://doi.org/10.1075/z.64.12par

Partington, A. (2006). *The linguistics of laughter: A corpus-assisted study of laughter-talk*. Routledge. https://doi.org/10.4324/9780203966570

Pérez-Paredes, P. (2019). The pedagogic advantage of teenage corpora for secondary school learners. *Data-Driven Learning for the Next Generation*, 67–87. https://doi.org/10.4324/9780429425899-5

Pérez-Paredes, P., & Alcaraz-Calero, J. M. (2009). Developing annotation solutions for online data driven learning. *ReCALL, 21*(1), 55-75. https://doi.org/10.1017/S0958344009000093

Perkuhn, R., Keibel, H., & Kupietz, M. (2012). *Korpuslinguistik*. Fink. https://doi.org/10.36198/9783838534336

Pharao, N., Maegaard, M., Møller, J. S., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society, 43*(1), 1–31. https://doi.org/10.1017/S0047404513000857

Pichler, H., & Levey, S. (2011). In search of grammaticalization in synchronic dialect data: General extenders in northeast England. *English Language and Linguistics, 15*, 441–71. https://doi.org/10.1017/S1360674311000128

Pilav, S. (2008). *Üniversite birinci sınıf öğrencilerinin söz varlığı üzerine bir araştırma* [Unpublished doctoral dissertation]. Gazi University.

Pomerantz, A. (1978). Complement responses: Notes on the co-operation of multiple constraints. In J. Schenkein (Ed.), *Studies in the Organization of Conversational Interaction* (pp. 79-112). Academic Press. https://doi.org/10.1016/B978-0-12-623550-0.50010-0

Powell, M. J. (1985). Purposesive vagueness: An evaluative dimension of vague quantifying expressions. *Journal of Linguistics, 21*, 31-50. https://doi.org/10.1017/S002222670001001X

Precht, K. (2008). Sex similarities and differences in stance in informal American conversation. *Journal of Sociolinguistics, 12*(1), 89-111. https://doi.org/10.1111/j.1467-9841.2008.00354.x

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *The International Journal of Bilingualism, 12*(1–2), 43–61. https://doi.org/10.1177/13670069080120010401

Quist, P. (2010). The sociolinguistic study of youth and multilingual practices in Denmark: An overview. In P. Quist, & B. A. Svendsen (Eds.), *Multilingual urban Scandinavia* (pp. 6–11). Multilingual Matters. https://doi.org/10.21832/9781847693143-005

Quist, P., & Svendsen, B. A. (2010). *Multilingual Urban Scandinavia: New linguistic practices.* Multilingual Matters. https://doi.org/10.21832/9781847693143

Rampton, B. (1995). *Crossing: Language and ethnicity among adolescents*. Longman.

Rampton, B. (1998). Language crossing and the redefinition of reality. In P. Auer (Ed.), *Codeswitching in conversation* (pp. 290–320). Routledge.

Rampton, B. (2006). *Language in late modernity: Interaction in an urban school*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486722

Rampton, B. (2011). From 'Multi-ethnic adolescent heteroglossia' to 'Contemporary urban vernaculars'. *Language & Communication, 31*(4), 276–94. https://doi.org/10.1016/j.langcom.2011.01.001

Rampton, B. (2013). From 'youth language' to contemporary urban vernaculars. In A. Deppermann (Ed.), *Das Deutsch der Migranten* (pp. 59–80). Mouton de Gruyter. https://doi.org/10.1515/9783110307894.59

Rampton, B. (2015). Contemporary urban vernaculars. In J. Nortier, & B. A. Svendsen (Eds.), *Language, youth and identity in the 21st century. Linguistic practices across urban spaces* (pp. 25–44). Cambridge University Press. https://doi.org/10.1017/CBO9781139061896.003

Rehbein, I., Schalowski, S., & Wiese, H. (2014). The KiezDeutsch Korpus (KiDKo) Release 1.0. *In Proceedings of the 9th International Conference on Language Resources and Evaluation* (pp. 367-375). European Language Resources Association.

Rehbein, J., Schmidt, T., Meyer, B., Watzke, F., & Herkenrath, A. (2004). *Handbuch für das computergestützte Transkribieren nach HIAT. In Arbeiten zur Mehrsprachigkeit, Folge B* (p. 56). http://www1.uni hamburg.de/exmaralda/files/azm_56.pdf

Rendle-Short, J. (2007). 'Catherine, you're wasting your time': Address terms within the Australian political interview. *Journal of Pragmatics, 39*(9), 1503-1525. https://doi.org/10.1016/j.pragma.2007.02.006

Rendle-Short, J. (2009). The address term mate in Australian English: Is it still a masculine term? *Australian Journal of Linguistics, 29*(2). 245–268. https://doi.org/10.1080/07268600902823110

Rendle-Short, J. (2010). 'Mate' as a term of address in ordinary interaction. *Journal of Pragmatics, 42*(5). 1201–18. https://doi.org/10.1016/j.pragma.2009.09.013

Reppen, R. (2010). Building a corpus: What are the key considerations In A. A. O'Keeffe & M. McCarty (Eds.), *The Routledge handbook of corpus linguistics* (pp. 38-52). Routledge. https://doi.org/10.4324/9780203856949.ch3

Rodríguez-González, F., & Stenström, A.-B. (2011). Expressive devices in the language of English and Spanish-speaking youth. *Revista Alicantina de Estudios Ingleses, 24*, 235–56. https://doi.org/10.14198/raei.2011.24.10

Roels, L., & Enghels, R. (2020). Age-based variation and patterns of recent language change: A case-study of morphological and lexical intensifiers in Spanish. *Journal of Pragmatics, 170*, 125-38. https://doi.org/10.1016/j.pragma.2020.08.017

Roels, L., De Latte, F., & Enghels, R. (2021). Monitoring 21st-Century real-time language change in Spanish youth speech. *Languages*, *6*(4), 162. https://doi.org/10.3390/languages6040162

Rørbeck Nørreby, T., & Spindler Møller, J. (2015). Ethnicity and social categorization in on- and offline interaction among Copenhagen adolescents. *Discourse, Context & Media, 8*, 46–54. https://doi.org/10.1016/j.dcm.2015.05.006

Roulet, E. (1980). Interactional markers in dialogue. *Applied Linguistics, 1*(3), 224-33. https://doi.org/10.1093/applin/1.3.224

Ruhi, Ş. (2009). The pragmatics of yani as a parenthetical marker in Turkish: Evidence from the METU Turkish Corpus. *Working Papers in Corpus-based Linguistics and Language Education, 3*, 285-98.

Ruhi, Ş. (2013). Interactional markers in Turkish: A corpus based perspective. *Journal of Linguistics and Literature*, *(10)*2, 1-7.

Ruhi, Ş., Hatipoğlu, Ç., Işık-Güler, H., & Eröz-Tuğa, B. (2010). *A guideline for transcribing conversations for the construction of Spoken Turkish Corpora using EXMARaLDA and HIAT*. Setmer Yayıncılık.

Rühleman, C. (2007). *Conversation in context: A corpus-driven approach*. Bloomsbury.

Rymes, B., & Leone, A. R. (2014). Citizen Sociolinguistics: A new media methodology for understanding language and social life. *Working Papers in Educational Linguistics, 29*(2), 25-43. https://repository.upenn.edu/wpel/vol29/iss2/4

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

Salakhyan, E. (2012). The Tübingen Corpus of Eastern European English (TCEEE): From a small-scale corpus study to a newly emerging non-native English variety. In S. Łodej & J. G. Newman (Eds.), *Token: A Journal of English Linguistics. Volume 1*. (pp. 143-57). Jan Kochanowski University Press.

Sargın, N. (1990). *Lise I ve lise III. sınıf öğrencilerinin durumluk-sürekli kaygı düzeylerinin belirlenip karşılaştırılması* [Unpublished master's thesis]. Dokuz Eylül University.

Say, B., Zeyrek, D., Oflazer, K., & Özge, U. (2004). Development of a corpus and a treebank for present-day written Turkish. In S. Yağcıoğlu (Ed.), *Proceedings of 12th International Conference on Turkish Linguistics* (pp. 183–92). Dokuz Eylul Publishing.

Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analysing discourse: Text and talk* (pp. 71–92). Georgetown University Press.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70*(6), 1075–95. https://doi.org/10.1525/aa.1968.70.6.02a00030

Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611841

Schiffrin, D., Tannen, D., & Hamilton, H. E. (2005). *The Handbook of Discourse Analysis.* Blackwell. https://doi.org/10.1002/9780470753460

Schmidt, T., & Wörner, K. (2014). EXMARaLDA. In T. Schmidt (Ed.), *Handbook on Corpus Phonology* (pp. 402-19). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199571932.013.030

Schwarz-Friesel, M., & Consten, M. (2011). Reference and anaphora. In W. Bublitz and N. R. Norrick (Eds.), *Foundations of pragmatics, vol. 1 of handbooks of pragmatics* (pp. 347–72). Mouton de Gruyter. https://doi.org/10.1515/9783110214260.347

Semerci, P. U., Erdoğan E., & Sandal Önal, E. (2017). *'BİZ'liğin aynasından yansıyanlar, Türkiye gençliğinde kimlikler ve ötekileştirme*. İstanbul Bilgi University Publishing.

Şengil, O. (1985). *Anadolu liseleri, özel Türk ve yabancı liselerine giriş sınavlarındaki tavan ile taban puanları ve cinsiyet ile başarı arasındaki ilişki* [Unpublished master's thesis]. Gazi University.

Sercombe, H. (2015). Youth in global/historical context. In M. Heathfield and D. Fusco (Eds.), Youth and Inequality in Education: Global Actions in Youth Work (pp. 17-32). Routhledge. https://doi.org/10.4324/9781315750668-2

Sezer, T., & Sezer, B. (2013). TS Corpus: Herkes için Türkçe derlem. *Proceedings* of the *27th National Linguistics Conference* (pp. 217-225). Hacettepe University, Turkey.

Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B. V., Krasny, M. E., & Bonney, R. (2012). Public participation in scientific research: A framework for deliberate design. *Ecology and Society, 17*(2). https://doi.org/10.5751/es-04705-170229

Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments. *Research on Language & Social Interaction, 39(*4), 377–409. https://doi.org/10.1207/s15327973rlsi3904_2

Sinclair J. McH., & Coulthard, R. M. (1975). *Towards an Analysis of discourse*. Oxford University Press.

Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp.410-30). Longmans.

Sinclair, J. (1992). Trust the text: The implications are daunting. In M. Davies & L. Ravelli (Eds.), *Advances in systemic linguistics* (pp. 5–19). Pinter.

Sinclair, J. (1996). *EAGLES preliminary recommendations on corpus typology.* Expert Advisory Group on Language Engineering Standards. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.1988&rep=rep1&type=pdf

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge. https://doi.org/10.4324/9780203594070

Sinclair, J. (2005). Corpus and text – Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxbow Books. http://ahds.ac.uk/linguistic- corpora/

Şişman, Ç. E. (2013). Üst-orta sınıf genç İstanbullular'ın gündelik dilindeki "Öteki" temsilleri: "Apaçi", "Kıro", "Amele", "Varoş", "Tiki", "Concon", "Ciks", In D. Lüküslü & H. Yücel (Eds.), *Gençlik Halleri: 2000'li Yıllar Türkiye'sinde genç olmak* (pp. 72-91). Efil Publishing.

Speer, S. A. (2012). The interactional organization of self-praise. *Social Psychology Quarterly, 75*(1), 52–79. https://doi.org/10.1177/0190272511432939

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.

Steingrímsson, S., Helgadóttir, S., Rögnvaldsson, K., Barkarson, S., & Guðnason, J. (2018). Risamálheild: A Very Large Icelandic Text Corpus. *In Proceedings of the 11th International Conference on Language Resources and Evaluation* (pp. 4361-66). European Language Resources Association.

Stenström, A.-B. (1997). Can I have a chips please? - Just tell me what one you want: Nonstandard grammatical features in London teenage talk. In J. Aarts (Ed.), *Studies in English language and teaching* (pp. 141-52). Rodopi.

Stenström, A.-B. (1998). From sentence to discourse: Cos (because) in teenage talk. In A. Jucker & Y. Ziv (Eds.), *Discourse markers: Descriptions and theory* (pp. 127-46). John Benjamins. https://doi.org/10.1075/pbns.57.08ste

Stenström, A.-B. (2005). It is very good eh– Está muy bien eh. Teenagers' use of tags – London and Madrid compared. In K. McCafferty, Tove Bull, & K. Killie (Eds.), *Contexts - historical, social, linguistic. Studies in celebration of Toril Swan* (pp. 279-91). Peter Lang.

Stenström, A.-B. (2007). The Spanish discourse markers o sea and pues and their English correspondences. In K. Aijmer & A.-M. Simon-Vandenbergen (Eds.), *Pragmatic markers in contrast*. Elsevier.

Strenström, A.-B. (2013). Youngspeak: Spanish vale and English okay. In K. Aijmer & B. Altenberg (Eds.), *Advances in Corpus-based Constrastive Linguistics* (pp. 127-139). John Benjamins. https://doi.org/10.1075/scl.54.08ste

Stenström, A.-B. (2014). *Teenage talk: From general characteristics to the use of pragmatic markers in a contrastive perspective.* Palgrave Macmillan. https://doi.org/10.1057/9781137430380_4

Stenström, A.-B. (2017). Swearing in English and Spanish teenage talk. In K. B. Fagersten and K. Stapleton (Eds.), *Advances in swearing research: New languages and new contexts* (pp. 157-79). John Benjamins. https://doi.org/10.1075/pbns.282.08ste

Stenström, A.-B. Andersen, G., & Hasund, K. (2002). *Trends in Teenage Talk*. John Benjamins. https://doi.org/10.1075/scl.8

Stenström, A.-B., & Jørgensen, A. M. (2008). A question of politeness? A contrastive study of phatic language in teenage conversation. *Pragmatics, 18*(4), 636-57.

Stenström, A.-B., Andersen, G., Hasund, K., Monstad K., & Aas, H. (1998). *User's Manual to Accompany The Bergen Corpus of London Teenage Language (COLT).* University of Bergen.

Stoffel, C. (1901). *Intensives and down-toners: A study in English adverbs.* Carl Winter's Universitätsbuchhandlung.

Stubbe, M., & Holmes, J. (1995). You know, eh, and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication, 15*(1), 63-88. https://doi.org/10.1016/0271-5309(94)00016-6

Stubbs, M. (1983). *Discourse analysis: The sociolinguistic analysis of natural language.* University of Chicago Press.

Svendsen, B. A. (2018). The dynamics of citizen sociolinguistics. *Journal of Sociolinguistics, 22*(2), 137–60. https://doi.org/10.1111/josl.12276

Svendsen, B. A., & Røyneland, U. (2008). Multiethnolectal facts and functions in Oslo, Norway. *International Journal of Bilingualism, 12(*1-2), 63–83. https://doi.org/10.1177/13670069080120010501

Syed, M. (2019). The Open Science Movement is For All of Us [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cteyb

Tagliamonte, S. (2006). 'So cool, right?': Canadian English entering the 21st century. Canadian English in a Global Context. *Special issue of Canadian Journal of Linguistics, 51*(2/3), 309–31. https://doi.org/10.1017/S0008413100004126

Tagliamonte, S. (2008). So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics, 12(*2), 361-94. https://doi.org/10.1017/S1360674308002669

Tagliamonte, S. (2016a). *Teen talk: The language of adolescents.* Cambridge University Press. https://doi.org/10.1017/CBO9781139583800

Tagliamonte, S. (2016b). So sick or so cool? The language of youth on the internet. *Language in Society, 45*(1), 1–32. https://doi.org/10.1017/S0047404515000780

Tagliamonte, S., & Roberts, R. (2005). So weird; so cool; so innovative: The use of intensifiers in the television series Friends. *American Speech, 80*(3), 280–300. https://doi.org/10.1215/00031283-80-3-280

Tagliamonte, S. A., & D'Arcy, A. (2007). Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change, 19*, 1–19. https://doi.org/10.1017/S095439450707007X

Tagliamonte, S. A., & D'Arcy, A. (2009). Peaks beyond phonology: Adolescence, Incrementation, and language change. *Language, 85*(1), 58–108. https://doi.org/10.1353/lan.0.0084

Tagliamonte, S. A., & Denis, D. (2010). The stuff of change: General extenders in Toronto, Canada. *Journal of English Linguistics, 38*(4), 335-68. https://doi.org/10.1177/0075424210367484

Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech, 83*(1), 3–34. https://doi.org/10.1215/00031283-2008-001

Tao, H., & Xiao, R. (2007). A corpus-based sociolinguistic study of amplifiers in British English. *Sociolinguistic Studies, 1*(2). 241–73. https://doi.org/10.1558/sols.v1i2.241

Tekin, M. (2015). Turkish discourse particles 'işte' and 'falan' and their English equivalents 'well' and 'like': A corpus study. *Journal of History School, 8*(21), 435-59. https://doi.org/10.14225/Joh673

Telli-Aydemir, A. (2009). Turkish National Report, In M. A. Andres, T. Olsson, A. B. Boado, & F. Miegel (Eds.,) *A qualitative analysis of European web based civic participation among young people* (pp. 127–45). University of London.

Temur, T. (2006). *İlköğretim 4. ve 5. sınıf öğrencilerinin yazı dilindeki kelime hazinelerinin bazı değişkenler açısından incelenmesi* [Unpublished doctoral dissertation]. Gazi University.

Tezcan, M. (1981). *Kuşaklar çatışması, okuyan ve çalışan gençlik üzerine bir araştırma.* Gül Yayınevi.

Thornborrow, J., & Morris, D. (2004). Gossip as strategy: The management of talk about others on reality TV show 'Big Brother'. *Journal of Sociolinguistics, 8*, 246–71. https://doi.org/10.1111/j.1467-9841.2004.00260.x

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Benjamins. https://doi.org/10.1075/scl.6

Toğrol, B. (2012). Qualifiers used in the daily language of adolescent Turkish boys. *Studies in Psychology, 7*(0), 82-88.

Tosunoğlu, M. (1998). *İlköğretim okuluna başlayan öğrencilerin okuma yazmayı öğrenmeden önceki kelime serveti üzerine bir araştırma* [Unpublished doctoral dissertation]. Gazi University.

Tottie, G. (1991). Conversational style in British and American English: The case of backchannels. In K. Aijmer & B. Altenberg. (Eds.), *English Corpus Linguistics* (pp. 254-71). Longman.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-49). Longman.

Trudgill, P. (1974). *The social differentiation of English in Norwich.* Cambridge University Press.

Trudgill, P. (1983). *Sociolinguistics: An introduction to language and society.* Penguin Books.

Tsakona, V., & Sifianou, M. (2019). Vocatives in service encounters: Evidence from Greek. *Acta Linguistica Hafniensia, 51*, 1-30. https://doi.org/10.1080/03740463.2019.1594578

TÜAD. (2012). *Ses 2012 Lansman Toplantısı.* TÜAD.

Tunçel, H. (2011). *Yabancılara Türkçe öğretiminde sözlü Türkçenin kelime sıklığı ve yaygınlığını belirleme çalışması* [Unpublished master's thesis]. Çanakkale Onsekiz Mart University.

Tütün, A. (1995). *Gençlik ve uyuşturucu madde alışkanlığı,* Milli Eğitim Bakanlığı Yayınları.

Tuzcu Tığlı, İ. (2019). *Neo-subcultures in the network society: The case of RKOT on Instagram* [Unpublished doctoral dissertation]. Galatasaray University.

Tüzün, S. (2000). Kentsel Türkiye hane ve bireyleri için bir tabakalaşma modeli olarak veri sosyo ekonomik statü indeksi (Veri S.E.S.İ). In F. Atacan, F. Ercan, H. Kurtuluş, & M. Türkay (Eds.), *Mübeccel Kıray içinyazılar* (pp. 371-85). Bağlam.

Uçar, A., & Yıldız, İ. (2015). Humour and impoliteness in Turkish: A corpus-based analysis of the television show Komedi Dükkanı 'Comedy Shop'. In Ş. Ruhi & Y. Aksan (Eds.), *Exploring (Im)politeness in specialized corpora: Converging methodologies and analytic procedures* (pp. 40-81). Cambridge Scholars Publishing.

Ünsal, H. (2005). *İlköğretim 5 ve 8. Sınıflar ile ortaöğretim 1.1 sınıfların sözlü anlatımdaki aktif kelime hazineleri: Afyon Merkez örneği* [Unpublished master's thesis]. Afyon Kocatepe University.

Üstüner, A. (2003). *Türkçede Pekiştirme*. Fırat Üniversitesi Basımevi.

Vaughan, E., McCarthy, M., & Clancy, B. (2017). Vague category markers as turn-final items in Irish English. *World Englishes, 36*(2), 208-23. https://doi.org/10.1111/weng.12254

Wilson, A., & Zeitlyn, D. (1995). The distribution of person-referring expressions in natural conversation. *Research on Language and Social Interaction, 28*(1), 61-92.

Wilson, N. (2010). Bros, boys and guys: Address term function and communities of practice in a New Zealand rugby team. *New Zealand English Journal, 24*, 33-54. https://doi.org/10.1207/s15327973rlsi2801_2

Woelfle, M., Olliaro, P., & Todd, M. H. (2011). Open science is a research accelerator. *Nature Chemistry, 3*(10), 745–48. https://doi.org/10.1038/nchem.1149

Wood, L., & R. Kroger (1991). Politeness and forms of address. *Journal of Language and Social Psychology, 10*(3), 145-68. https://doi.org/10.1177/0261927X91103001

Yahşi, Ö. (2020). *Determination of basic level vocabulary in teaching Turkish to foreigners: Vocabulary in Turkish textbooks prepared for foreigners with vocabulary in written expression and spoken expression of learners of Turkish as a foreign language* [Unpublished doctoral dissertation]. Sakarya University.

Yaman, Ö. M. (2010). *Türkiye gençlik araştırmaları bibliyografyası 1923-2010*. Gençlik ve Spor Genel Müdürlüğü Gençlik Hizmetleri Daire Başkanlığı Kültür Yayınları.

Yaman, Ö. M. (2013). A bibliographic assessment on studies of sociology of youth in Turkey. *Alternatif Politika, 5*(2), 114-38.

Yang, H., & Wei, N. (2005). *Construction and data analysis of a Chinese learner spoken English corpus.* Shanghai Foreign Language Education Press.

Yazı, Z. (2005). *Orta öğretim 10. sınıf öğrencilerinin kelime hazinesi üzerine bir araştırma* [Unpublished master's thesis]. Gazi University.

Yentürk, N. & Başlevent, C. (2009) Türkiye'de genç işsizliği: Etkileyen faktörler ve politika önerileri, In N. Yenitürk, Y. Kurtalan, & G. Nemutlu (Eds.), *Türkiye'de gençlik çalışması ve politikaları* (pp. 345-79). Bilgi University Publishing.

Yılmaz, E. (1994). *Descriptive and comparative study of the discourse markers well in English and şey in Turkish* [Unpublished master's thesis]. University of Essex.

Yılmaz, E. (2004). *A pragmatic analysis of Turkish discourse particles: Yani, işte and şey* [Unpublished doctoral dissertation]. Middle East Technical University.

Yngve, V. H. (1970). On getting a word in edgewise. In M. A. Campbell (Ed.), *Papers from the Sixth Regional Meeting of Chicago Linguistic Society* (pp. 567-77). Chicago Linguistic Society.

Yolcu Yavuz, S. (2019). *Cumhuriyet Misyonerleri.* Vakıfbank Kültür Yayınları.

Yolcu, S. (2014). *The construction of youth as a political subject in Turkey* (1930-1946) [Unpublished doctoral dissertation]. Istanbul University.

Yonucu, D. (2005). *From the place of the 'Dangerous Classes' to the place of danger: Emergence of new youth subjectivities in Zeytinburnu* [Unpublished master's thesis]. Boğaziçi University.

Yörükoğlu, A. (1968). *Çocukların sosyal ihtiyaçları ve eğitimde ruh sağlığı*. Yeni Gün Matbaası.

Yücel, H., & Lüküslü, D. (2013). 2000'li yılları gençlik üzerinden okumak, In D. Lüküslü & H. Yücel (Eds.), *Gençlik halleri: 2000'li yıllar Türkiye'sinde genç olmak* (pp. 9-23). Efil Publishing.

Yüceol Özezen, M. (2013).Türkiye Türkçesinde pekiştirme işlevli Ki. *Journal of Turkish Studies, 8*(9), 367-73. https://doi.org/10.7827/TurkishStudies.5374

Yüceol Özezen, M. (2019). Türkçede seni/sizi seslenme sözleri ve Türkçede seslenme (vokatif) durumuna tarihsel bir bakış. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi, 8*(2), 726-742. https://doi.org/10.7884/teke.4521

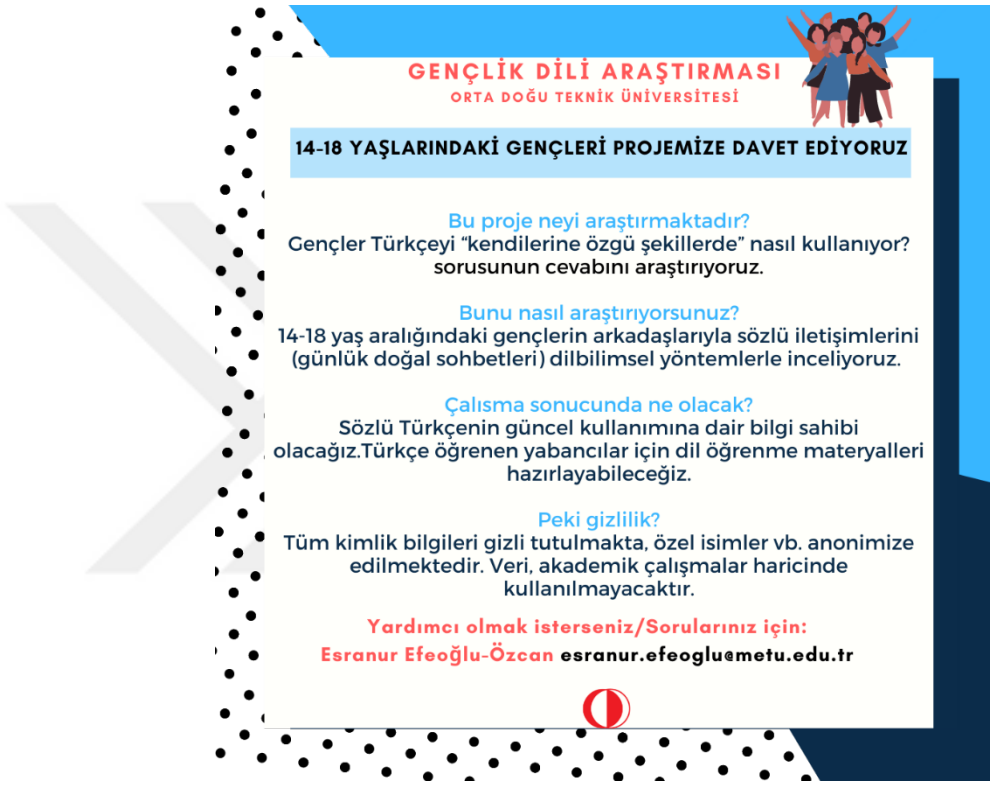
Zeyrek D., Demirşahin I., Sevdik-Çallı, A., & Çakıcı, R. (2013). Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue Discourse, 4*(2),174–84. https://doi.org/10.5087/dad.2013.208

Zeyrek, D. (2001). Politeness in Turkish and its linguistic manifestations: A sociocultural perspective. In A. Bayraktaroğlu & M. Sifianou (Eds.), *Linguistic politeness across boundaries: The case of Greek and Turkish* (pp. 43-74). John Benjamins. https://doi.org/10.1075/pbns.88.04zey

Zwicky, A. (1974). Hey, Whatsyourname! In M. W. La Galy, R. A. Fox & A. Bruck (Eds.), *Papers from the tenth regional meeting of the Chicago Linguistic Society* (pp. 787–801). Chicago Linguistic Society.

APPENDICES

A. ONLINE FLYER



YOUTH LANGUAGE PROJECT
WE INVITE YOUNG PEOPLE AGED 14-18 TO OUR PROJECT!

What is this project investigating?
We are investigating how young people use Turkish "in their own unique ways".

How are you researching this?
We use linguistic methods to analyze the verbal communication (everyday natural conversations) of young people between the ages of 14-18 among their friends.

What are the expected results of the study?
We will obtain information about the contemporary use of spoken Turkish. We will be able to prepare language learning materials for foreigners learning Turkish.

What about confidentiality?
All personal information is kept confidential, private names etc. are anonymized. The data will only be used for scientific purposes.

If you want to contribute/ask questions, please contact: esranur.efeoglu@metu.edu.tr

B. INFORMED CONSENT FORMS

B-1 Informed Consent Form for Parents/Guardians

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

Sayın Veliler, Sevgili Anne-Babalar,
Bu çalışma, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü İngiliz Dili Öğretimi doktora programında doktora öğrencisi Esranur Efeoğlu-Özcan'ın Doç. Dr. Hale Işık-Güler danışmanlığında yürütmekte olduğu doktora tez araştırmasıdır. Bu araştırma, ODTÜ İnsan Araştırmaları Etik Kurulu tarafından 150-ODTÜ-2019 protokol numarası ile onaylanmıştır.

✶**Bu projenin amacı nedir?**
Bu araştırmanın amacı, sözlü Türkçe gençlik dilinin söz varlığı ve gençlerin akranlarıyla iletişim dinamiklerinin incelenmesidir.

✶**Sizin ve çocuğunuzun katılımcı olarak ne yapmasını istiyoruz?**
Araştırmanın amacı kapsamında ve çocuklarınızın bu araştırmaya katılmasına izin verdiğiniz takdirde, çocuklarınızın akranlarıyla gerçekleştirdiği sohbetlerin ses kayıtlarına ihtiyaç duymaktayız. Sizden çocuğunuzun katılımcı olmasıyla ilgili izin istediğimiz gibi, çalışmaya başlamadan çocuğunuzdan da katılımıyla ilgili rızası mutlaka alınacaktır.

✶**Çocuğunuzdan alınan bilgi ve kayıtlar ne amaçla ve nasıl kullanılacak?**
Kayıtlarla elde edilecek veri ise yazıya çevrilecek ve bu araştırma kapsamında elde edilen sonuçlar yalnızca bilimsel yayınlarda kullanılacaktır. Çocuğunuzun ve sizin kimlik bilgileriniz ve kişisel bilgileriniz kesinlikle gizli tutulacak, anonimize edilecektir.

✶**Çocuğunuz ya da siz çalışmayı yarıda kesmek isterseniz ne yapmalısınız?**
Bu çalışmanın çocuğunuzun psikolojik gelişimine olumsuz etkisi olmayacağından emin olabilirsiniz. Yine de, bu formu imzaladıktan sonra çocuğunuz katılımcılıktan ve bu araştırmadan ayrılma hakkına sahiptir. Çalışma süresince herhangi bir nedenden ötürü rahatsızlık hisseden katılımcılar araştırmadan ayrılabilir, böyle bir durumda araştırmacıyla iletişime geçmeniz yeterli olacaktır.

Bu çalışmayla ilgili daha fazla bilgi almak isterseniz:

Esranur Efeoğlu-Özcan

Size ulaşan proje katılımcısının (ses kaydını toplayan kişi) adı soyadı: ……..
Çocuğunuzun adı ve soyadı : ……
Adınız ve soyadınız : ……..
İletişim bilgileriniz (e-posta veya telefon) : ……

Çalışmaya gönüllü katılımınızı onaylıyorsanız, lütfen aşağıdaki kutucuğu işaretleyiniz.

☐ **Bu araştırmaya tamamen gönüllü olarak katılıyorum ve yukarıda adı geçen velisi/vasisi olduğum katılımcının da bu araştırmada katılımcı olmasına izin veriyorum. Verdiğim bilgilerin bilimsel amaçlı olarak kullanılmasını kabul ediyorum.**

B-2 Informed Consent Form for Participants

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

Sayın Katılımcı,

Bu çalışma, Orta Doğu Teknik Üniversitesi, Eğitim Fakültesi, Yabancı Diller Eğitimi Bölümü'nden Doç. Dr. Hale Işık-Güler ve Esranur Efeoğlu-Özcan'ın birlikte yürütmekte olduğu "Gençlik Dili" Projesinin bir parçasıdır.

Bu araştırma projesi, ODTÜ İnsan Araştırmaları Etik Kurulu tarafından 150-ODTÜ-2019 protokol numarası ile onaylanmıştır.

✶Bu projenin amacı nedir?
Bu araştırmanın amacı, sözlü Türkçe gençlik dilinin söz varlığı ve gençlerin akranlarıyla iletişim dinamiklerinin incelenmesidir.

✶Sizin katılımcı olarak ne yapmanızı istiyoruz?
Araştırmanın amacı kapsamında akranlarınızla sohbetlerin kayıtlarına ihtiyaç duymaktayız. Kişisel bilgileriniz kesinlikle gizli tutulacak ve sadece bilimsel araştırma amacıyla kullanılacaktır. İsim ve kimlik bilgileriniz, hiçbir şekilde kimseyle paylaşılmayacaktır. Kayıtlarla elde edilecek veri ise yazıya çevrilecek ve elde edilen sonuçlar yalnızca bilimsel yayınlarda kullanılacaktır.

Bu çalışmayla ilgili daha fazla bilgi almak isterseniz (xxx) eposta adresinden ya da telefon aracılığı ile (xxx) bize ulaşabilirsiniz.

Katkılarınız ve izniniz için teşekkür ederiz.

Saygılarımızla,

Esranur Efeoğlu-Özcan

Size ulaşan proje katılımcısının (ses kaydını toplayan kişi) adı soyadı: .....
Adınız ve soyadınız : .....
İletişim bilgileriniz (e-posta veya telefon) : .....

Çalışmaya gönüllü katılımınızı onaylıyorsanız, lütfen aşağıdaki kutucuğu işaretleyiniz.

☐ **Bu araştırmaya tamamen gönüllü olarak katılıyorum ve verdiğim bilgilerin bilimsel amaçlı olarak kullanılmasını kabul ediyorum.**

C. DATA COLLECTION TIMELINE

Year	*Month*	*Local & Global Events*	*Type of Interaction*	*Total N of recordings*	*Duration of recordings*
2019	October	Ongoing 2019-2020 Fall Semester.	Face to Face (*n*=2)	6	1 hr 55 min
	November	News reports regarding cases of COVID-19 infections in Wuhan, China.	Face to Face (*n*=4)		
	December	First case of COVID-19 officially documented.	-		
2020	January	Coronavirus Scientific Board was established byTurkish government.	-	27	14 hr 56 min
	February	Social media and mainstream media campaigns started to disseminate precautions for avoiding virus transmission. Travel restrictions for China and Iran were implemented.	Online (*n*=1)		
	March	First case of COVID-19 in Turkey was officially reported on 10.03 2020. World Health Organization dclared the ongoing outbreak as a global pandemic. First death due to COVID-19 in Turkey was officially reported on 14.03.2020 All schools were closed starting on 16.03.2020. Distance education started on 23.03.2020, mainly via (EBA). Cinemas, cafes & restaurants, gyms, concert halls, mosques, malls were closed. All sports events as well as scientific and cultural meetings were cancelled.	-		
	April	Government imposed partial curfew for those under the age of 20. Curfew for all citizens was implemented on the weekend of April 11th. Borders of 31 provinces were shut down except for transiting essential supplies.	-		
	May	It was announced that schools would remain closed until the end of May. People aged 15-20 was allowed to go outside on May 15th, from 11:00 to 15:00.	-		
	June	A roadmap for normalization period was announced, curfew restrictions were eased. National University/High School Entrance Exams were held face-to-face.	-		
	July	Tourism activities were resumed.	-		
	August	Due to the resurgence of infections, reopening of schools was postponed.	-		

(cont'd)

Year	*Month*	*Local & Global Events*	*Type of Interaction*	*Total N of recordings*	*Duration of recordings*
2020	September	Kindergarten and first year students resumed education without mandatory attendance.	-		
	October	Government responded to the speculations concerning the number of COVID-19 cases.	-		
	November	Cerfew on people who are above 65 years and people younger than 20 years is reinstated.	Face to Face (*n*=2) & Online (*n*=3)		
	December	The number of daily deaths reached a peak in the country.	Face to Face (*n*=7) & Online (*n*=14)		
2021	January	COVID-19 vaccines were started to be administered to people. 8th and 12th grade students started attending face-to-face classes at private cram schools to study for high school and university entrance exams.	-	16	9 hr 20 min
	February	Preparations for face-to-face education started. Nation-wide administration of vaccines to teachers.	-		
	March	Restrictions were eased by the government. In-class education started based on the assessment of local risk levels of provinces.	-		
	April	Due to the infection rate, nationwide lockdown was implemented.	Online (*n*=1)		
	May	Curfew restrictions for people below 18 year olds were dropped.	Face to Face (*n*=2) & Online (*n*=10)		
	June	High schoolers returned to school to attend in-class education with masks on.	Online (n=1)		
	July	Curfew restrictions were dropped completely, restaurants resumed activities without restrictions.	Face to Face (n=1)		
	August	Age for eligibility for vaccination was lowered to 15 years.	-		
	September	All levels of students resumed in-class face-to-face education.	-		
	October	The number infections increased in schools, classes were put in quarantine if a student was diagnosed with COVID-19.	Face to Face (n=1)		
Total	***49 recordings: Face to Face (n=19)& Online (n=30), 26 hr 11 minutes of data***				

D. SEMI-STRUCTURED INTERVIEW PROTOCOL

Date : ……..

Time : ……..

Mode of interview : ☐ Face-to-Face ☐ Online: ……..

Interviewee Name-Surname : ……..

Thank you for taking the time to participate in this interview. The purpose of this interview is to obtain information about daily and online routines, activity types, personal interests and hobbies of young people between ages 14 to 18 in Turkey. The information gathered will be used for developing the content of the questionnaire to be used for The Corpus of Turkish Youth Language Project.

In this interview, I will present a couple of guiding questions to you. There is no correct answer to any of these questions. You can provide answers as long as you like. If you need additional explanation or wish to learn more about the justification behind any questions, do not hesitate to ask for clarification. You can expand on your answers if you feel it would be helpful for this study.

The interview will approximately take 15 to 20 minutes. I will record the audio of this talk and I will be taking notes while we are talking. All of your personal information will be kept confidential and anonymous, your answers will only be used for research purposes.

The prompts below are the guiding themes for researcher to use in the interview:

- **Demographic details:** name-surname, age, grade, city of residence, school type
- **Daily routines & activities:** the places the interviewee like to go in their free time, activities there, with whom, the frequency of visits; daily routines at home/weekeds/vacations
- **Online activities**: social media, websites, applications used and the frequency of use
- **Friends**: profile of close friends and extended friend groups, frequency of interaction, modes of interaction
- **Hobbies & Interest:** personal hobbies and interests, salient hobbies and interests of interviewee's peer groups
- **School**: favourite subject, least favourite subject, work load, the interaction with teachers, the interaction with peers during break times, routines of commuting to school, school related extracurricular activities if there are any
- **Entertainment**: tv/online shows, series, movies; favourite actors/actresses (personal vs peer favourites)
- **Music:** music taste, favourite artists (personal vs peer favourites)
- **People**: authors, tv personalities, influencers, etc. (personal vs peer favourites)
- **The future**: dreams and goals concerning future (academic, personal, etc)
- **Family**: interaction with family (members of family, occupations of parents)

E. RECORDING LOG

This is the body of text used for online questionnaire. The participants access to the form via the link provided by the researcher, fill and send the form online. The questions required answers as text or multiple-choice selection. All the questions were compulsory except for those in section 4 and 7.

[Abridged Version] Within the scope of this project;
We examine the natural conversations of Turkish speaking young people between the ages of 14-18 in Turkey. This project is conducted by Esranur Efeoğlu-Özcan within the scope of her doctoral dissertation supervised by Assoc. Prof. Dr. Hale Işık-Güler from Middle East Technical University. This research has been approved by METU Human Research Ethics Committee with protocol number 150-ODTÜ-2019. Please fill this form and the consent forms after you completed your recordings.

SECTION 1 : RECORDING

Provide information about the person who recorded the audio and filled in this form along with information regarding the audio file. / *Ses kaydını yapan ve bu formu dolduran kişi hakkında bilgi vererek, ses dosyasına ilişkin bilgileri yazınız.*

- Name-Surname */ Adınız-Soyadınız*
- Date of Recording */ Kayıt tarihi*
- Length of Recording (in minutes : seconds) */ Kayıt uzunluğu (dakika : saniye)*
- Name of the audio file */ Ses kaydının ismi*

SECTION 2 : SPEAKER INFO

(Sections 3 and 4 are duplicates of this section, section 3 is compulsory while section 4 is optional) Provide detailed information on all the speakers in this recording, one by one. You will fill a separate section for each of the speakers in the recording. For the information requested, contact the speakers yourself, or through a friend/relative and ask them to provide you with the info below. / *Bu kayıttaki her bir konuşmacı için ayrı bir bölüm doldurarak tüm konuşmacılar hakkında tek tek ayrıntılı bilgi veriniz. İstenen bilgiler için konuşmacılarla kendiniz veya bir arkadaşınız/akrabanız aracılığıyla iletişime geçiniz ve aşağıdaki bilgileri size vermelerini isteyiniz.*

- Name-Surname / *Ad-Soyad*
- Date of birth / *Doğum tarihi*
- Sex:
  - ☐ Male / *Erkek*
  - ☐ Female / *Kız*
  - ☐ Prefer not to Say / *Belirtmek İstemiyorum*
- City they currently live in / *Şu anda yaşadığı şehir*
- Hometown / *Memleketi*
- Mother tounge / *Anadili*

- Other languages known / *Bildiği diğer diller*
- Level of education / *Eğitim seviyesi*
    - ☐ High School / *Lise*
    - ☐ High School Graduate / *Lise mezunu (henüz üniversiteye başlamamış)*
- Name of school / *Okulunun ismi*
- Grade level / *Sınıfı*
- GPA / *Haziran 20XX itibariyle yıl sonu başarı puanı*
- Mother's level of education / *Annesinin eğitim bilgisi*
    - ☐ Illiterate / *okur-yazar değil*
    - ☐ Primary or secondary school graduate / *ilkokul veya ortaokul mezunu*
    - ☐ High school graduate / *lise mezunu*
    - ☐ University graduate or above / *üniversite veya üstü*
- Mother's occupation / *Annesinin mesleği*
- Father's level of education / *Babasının eğitim bilgisi*
    - ☐ Illiterate / *okur-yazar değil*
    - ☐ Primary or secondary school graduate / *ilkokul veya ortaokul mezunu*
    - ☐ High school graduate / *lise mezunu*
    - ☐ University graduate or above / *üniversite veya üstü*
- Father's occupation / *Babasının mesleği*
- Number of siblings & ages / *Kardeş sayısı ve yaşları*
- For purposes of speaker identification: Please write this speaker's first utterances in the recording (to identify/not to confuse him/her) / *Kayıtta saptayabilmemiz için: bu konuşmacının ilk söylediği kelime/ifade nedir?*

- For purposes of speaker identification : Please write something (a line) only this speaker said as well as another feature (voice quality) that can help us identify the speaker / *Konuşmacıyı tanımlama amacıyla: Lütfen sadece bu konuşmacının söylediği bir şeyi (bir satır) ve konuşmacıyı tanımlamamıza yardımcı olabilecek başka sesi/konuşma şekli ile ilgili bir özelliği yazınız.*

SECTION 5 : SETTING

In this section we ask you to provide information about where and how the conversation took place. / *Bu bölümde konuşmanın nerede ve ne şekilde geçtiği hakkında bilgi vermenizi rica ediyoruz.*

- If all speakers were in the same place, describe their surroundings (furniture, mood, noise level, etc.). If the conversation took place online, describe where (their room, outside, etc.) each of the speakers participated in the chat. / *Konuşma nerede gerçekleşti? (Yüzyüze ise bulundukları ortamı/mekanı tasvir ediizn, online görüşme gerçekleştirildi ise hangi platformda kayıt alındığını ve konuşucuların konuşmaya nereden bağlandıklarını yazınız örn. kendi odaları)*

- What was the time of day? / *Kayıt saati nedir?*

SECTION 6 : RELATIONSHIP BETWEEN SPEAKERS

Describe the relationship between the speakers in detail below. / *Konuşucular arasındaki ilişki hakkında aşağıda bilgi veriniz.*

- Please indicate how the speakers met and how long the speakers had known each other and what their relationship was like prior to the conversation (are they classmates? close friends? etc.) / *Konuşucular nasıl tanıştı, ne zamandır birbirlerini tanıyorlar belirtiniz ve kayıt öncesinde aralarındaki ilişkiyi hakkında bilgi veriniz. (Sınıf arkadaşı mı, yakın arkadaş mı, vb belirtiniz.)*

- Please indicate how often these speakers communicate face-to-face or through online communication channels. / *Konuşucuların yüzyüze veya uzaktan iletişim araçları ile ne sıklıkla iletişim kurduğunu belirtiniz.*

  ☐ Every day / *Hergün*
  ☐ Every other day / *Her iki günde bir*
  ☐ Every 3-4 days / *Her üç dört günde bir*
  ☐ Once a week / *Haftada bir*
  ☐ Every other week / *İki haftada bir*
  ☐ Once a month / *Ayda bir*
  ☐ Less frequently / *Daha az sıklıkla*
  ☐ Never before / *Daha önce hiç görüşmediler*

SECTION 7 : ADDITIONAL COMMENTS
Overall, is there anything that you found interesting about the conversation or the speakers you would like to comment on? / *Genel anlamda bu kayıtla ilgili belirtmek istediğiniz bir husus var ise lütfen belirtiniz.*

CONSENT / *ONAY*
Within the scope of the study, your personal information will be kept confidential and will only be used for scientific research purposes. The name, surname and other personal information of you and other participants, as well as the identity information of all the persons mentioned in the recordings, will be anonymized. If you consent to your voluntary participation in the study, please check the box below.

☐ I am participating in this study completely voluntarily and I agree that the information I provide will be used for scientific purposes.

Çalışma kapsamında kişisel bilgileriniz kesinlikle gizli tutulacak ve sadece bilimsel araştırma amacıyla kullanılacaktır. Siz ve tüm katılımcıların isim-soyisim ve kimlik bilgileri ve ayrıca kayıtlarda sohbet sırasında bahsi geçen tüm şahısların kimlik bilgileri anonimize edilmektedir. Çalışmaya gönüllü katılımınızı onaylıyorsanız, lütfen aşağıdaki kutucuğu işaretleyiniz.

☐ *Bu araştırmaya tamamen gönüllü olarak katılıyorum ve verdiğim bilgilerin bilimsel amaçlı olarak kullanılmasını kabul ediyorum.*

F. TRANSCRIPTION CONVENTIONS

Symbol	*Function*	*Example for Turkish*	*Example for English*
•	micro pause	*çalışmamı okuduğunuz için • gönülden teşekkürler.*	*It takes a long time • to become young.*
((_._))	timed pause	*çalışmamı okuduğunuz için teşekkürler.* ***((1.0))*** *gönülden bir teşekkür bu.*	*It takes a long time to become young.* ***((1.0))*** *don't you think so?*
/	repair	*çalışmamın/ çalışmamı okuduğunuz için teşekkürler.*	*It takes a long to/ time to become young.*
.	falling intonation	*bu çalışma tamamlandı.*	*It takes a long time to become young.*
?	questions	*bu çalışma tamamlandı mı****?***	*Does it take a long time to become young?*
!	rising intonation	*bu çalışma tamamlansın****!***	*It took such a long time to become young!*
...	cut-off/ interrupted utterance	*bu çalışma...*	*It takes a really long time to...*
◡	latching	*aynen.* ◡*aynen. bitti.*	*It takes a long time.* ◡*becoming young.*
-	multi-syllable non-lexicalised units	*ı-ıh˙ bitmedi.*	*a-ha! it takes a long time to become young!*
˙	non-lexicalised units and paralinguistic features	*hıı˙ sanırım bu çalışma bitmek üzere.*	*eer˙ become young sure takes time.*

Symbol	*Function*	*Example for Turkish*	*Example for English*
((…))	non-linguistic features	*bu çalışmanın bitmesini istemiyorum* ***((laughs))***	*Never knew becoming young would take me this much time* ***((laughs))***
(text)	uncertain parts	*bu çalışma* ***(bitmek)*** *üzere.*	*It takes a* ***(long)*** *time to become young.*
((XXX))	unintelligible/ inaudible parts	*bu çalışma var ya* ***((XXX))****!*	*It takes a long time to* ***((XXX))****.*
<text>	overlaps (markup only in txt file)	*A: bu çalışma bitse* ***<****rahatlar mıyım?* ***>1>*** *B:* ***<****inşallah abi.* ***>1>*** *A:* ***<****hayda!* ***>2>*** *B:* ***<****rahatlarsın* ***>2>****rahatlarsın!*	*A: how do you define* ***<****the youth?* ***>1>*** *B:* ***<****just a word.* ***>1>*** *A:* ***<****a what?* ***>2>*** *B:* ***<****it is just a word****>2>*** *a mere word!*

G. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA/ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
[illegible]
[illegible]

Sayı: 28620816 / 164

26 MART 2019

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç.Dr. Hale Işık GÜLER

Danışmanlığını yaptığınız Esranur **EFEOĞLU'nun** "**Türk Gençlerinin Güncel Üretici Söz Varlığı Derleminin Oluşturulması ve Genç Dilinin Söylemsel Dinamiklerinin Belirlenmesi Projesi**" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **150-ODTÜ-2019** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız

Prof. Dr. Tülin GENÇÖZ
Başkan

Prof. Dr. Ayhan SOL
Üye

Prof. Dr. Ayhan Görbüz DEMİR
Üye

Prof. Dr. Yaşar KONDAKÇI
Üye

Doç. Dr. Emre SELÇUK
Üye

Doç. Dr. Pınar KAYGAN
Üye

Dr. Öğr. Üyesi Ali Emre TURGUT
Üye

H. THE PARTICIPANT SAMPLE

H-1: The distribution of participants by their hometown tabulated by provinces and regions of Turkey (Nomenclature of Territorial Units for Statistics)

Country	*Region*	*Province (by hometown)*	*No. of speakers*
TURKEY	TR1 Istanbul	TR100 İstanbul	15
		Total	**15**
	TR2 West Marmara	TR222 Çanakkale	5
		TR213 Kırklareli	1
		TR221 Balıkesir	3
		TR211 Tekirdağ	2
		Total	**11**
	TR3 Aegean	TR310 İzmir	8
		TR332 Afyonkarahisar	4
		TR322 Denizli	3
		TR323 Muğla	5
		TR321 Aydın	1
		TR333 Kütahya	3
		Total	**24**
	TR4 East Marmara	TR421 Koceli	4
		TR412 Eskişehir	3
		TR411 Bursa	1
		TR422 Sakarya	1
		Total	**9**
	TR5 West Anatolia	TR511 Ankara	10
		TR521 Konya	6
		Total	**16**
	TR6 Mediteranean	TR622 Mersin	5
		TR621 Adana	4
		TR611 Antalya	2
		TR632 Kahramanmaraş	3
		Total	**14**
	TR7 Central Anatolia	TR711 Kırıkkale	1
		TR721 Kayseri	3
		TR715 Kırşehir	2
		TR722 Sivas	2
		TR714 Nevşehir	1
		Total	**9**
	TR8 West Black Sea	TR821 Kastamonu	3
		TR831 Samsun	1
		TR834 Amasya	1

(cont'd)

Country	*Region*	*Province (by hometown)*	*No. of speakers*
		TR822 Çankırı	1
		TR811 Zonguldak	1
		Total	**7**
	TR9 East Black Sea	TR901 Trabzon	2
		TR905 Artvin	1
		Total	**3**
	TRA Northeast Anatolia	TRA11 Erzurum	1
		TRA22 Ağrı	1
		Total	**2**
	TRB Central East Anatolia	TRB12 Elazığ	2
		TRB14 Tunceli	1
		Total	**3**
	TRC Southeast Anatolia	TC11 Gaziantep	2
		TRC31 Mardin	1
		Total	**3**
		Missing	7
		Speakers All	**123**

H-2: The distribution of participants by their city of residence tabulated by provinces and regions of Turkey (Nomenclature of Territorial Units for Statistics)

Country	*Region*	*Province (by city of residence)*	*No. of speakers*
TURKEY	TR1 Istanbul	TR100 İstanbul	19
		Total	**19**
	TR2 West Marmara	TR222 Çanakkale	5
		TR213 Kırklareli	2
		TR221 Balıkesir	1
		Total	**8**
	TR3 Aegean	TR310 İzmir	8
		TR332 Afyonkarahisar	3
		TR322 Denizli	3
		TR323 Muğla	5
		TR321 Aydın	3
		TR333 Kütahya	2
		Total	**24**
	TR4 East Marmara	TR421 Koceli	6
		TR412 Eskişehir	5
		TR411 Bursa	3
		TR422 Sakarya	3
		Total	**17**
	TR5 West Anatolia	TR511 Ankara	22
		TR521 Konya	4

(cont'd)

Country	*Region*	*Province (by city of residence)*	*No. of speakers*
		Total	**26**
	TR6 Mediteranean	TR622 Mersin	5
		TR621 Adana	3
		TR611 Antalya	3
		TR632 Kahramanmaraş	1
		Total	**12**
	TR7 Central Anatolia	TR711 Kırıkkale	2
		Total	**2**
	TR8 West Black Sea	TR821 Kastamonu	3
		TR831 Samsun	2
		Total	**5**
	TR9 East Black Sea	TR901 Trabzon	1
		Total	**1**
	TRA Northeast Anatolia	TRA11 Erzurum	0
		Total	**0**
	TRB Central East Anatolia	TRB12 Elazığ	2
		Total	**2**
	TRC Southeast Anatolia	TC11 Gaziantep	0
		Total	**0**
		Missing	7
		Speakers All	**123**

I. VOCATIVES WITH GENDERED REFERENTS

I-1: Vocatives with gendered (male) referents tabulated by addressees and addressers in the corpus

Type	*Gloss*	*Addresser*		*Addressee*	
abi	*elder brother*	F	26	F	24
				M	2
		M	33	F	3
				M	33
aga	*elder brother*	F	3	F	2
				M	1
		M	8	F	2
				M	6
baba	*father*	F	2	F	2
				M	0
		M	3	F	0
				M	3
beyler	*gentlemen*	F	0	F	0
				M	0
		M	4	F	1
				M	3
birader	*sibling*	F	0	F	0
				M	0
		M	3	F	1
				M	2
bro	*brother*	F	2	F	2
				M	0
		M	3	F	0
				M	3
oğlum	*my son*	F	10	F	8
				M	2
		M	48	F	7
				M	41
pezevenk	*pimp*	F	1	F	1
				M	0
		M	2	F	0
				M	2

I-2: Vocatives with gendered (female) referents tabulated by addressees and addressers in the corpus

Type	*Gloss*	*Addresser*		*Addressee*	
anam	*my mother*	F	2	F	2
				M	0
		M	0	F	0
				M	0
hatun	*woman*	F	0	F	0
				M	0
		M	1	F	1
				M	0
kız	*girl*	F	2	F	2
				M	0
		M	0	F	0
				M	0

J. CHI-SQUARE TEST RESULTS

Intensifiers tabulated by tokens produced by female and male speakers

Rank	*Type*	*Gloss*	*AF by speaker*		*p*-value
			Female	*Male*	
1	çok	*very*	1200	505	<.000
2	biraz(cık)	*barely*	144	52	<.000
3	bayağı	*excessively*	115	73	<.000
4	en	*the most*	82	66	<.000
5	fazla	*excessively*	82	38	N.S.
6	gerçekten	*really*	90	22	<.000
7	aşırı	*excessively*	73	36	<.000
8	cidden	*seriously*	42	6	<.000
9	gayet	*excessively*	17	20	<.000
10	full	*full*	18	15	<.000
11	bir tık	*a bit*	21	5	<.000
12	kesinlikle	*absolutely*	14	3	<.000
13	valla(hi)	*really*	9	6	<.000
14	harbi(den)	*really*	3	11	<.000
15	özellikle	*particularly*	9	5	<.000
16	iyice	*quite*	10	3	<.000
17	iyi	*well*	8	4	<.000
18	ana +	*mother-plus swearing exp.*	1	8	<.000
19	öyle	*so*	3	5	<.000
20	manyak	*crazy*	4	1	<.000

AF: absolute frequency

K. CONCORDANCES FOR *AŞIRI* AND *BAYAĞI*

File	*Speaker*	*L*	*Node*	*R*	*Oriented towards*
Y-3-2M1F-01112019	SM09006	bişey yatıyo. çok mühendisçe.	**aşırı**	iyi.	*other*
Y-2-F-03122020-2	SF11003	aynen NineNine	**aşırı**	iyi.	*other*
Y-2-F-03122020-2	SF11003	müthiş!	**aşırı**	iyi.	*other*
Y-2-F-03122020-2	SF11003	aynen. müthiş ya!	**aşırı**	iyi.	*other*
Y-3-2M1F-09042021	SM10005	çok iyisin ((name_female)). sen	**aşırı**	iyisin ya zaten.	*interlocutor*
Y-2-F-02122020	SF09004	edebiyat dilbilgim çok iyi.	**aşırı**	iyi yani. edebiyat dilbilgisine	*self*

File	*Speaker*	*L*	*Node*	*R*	*Oriented towards*
Y-2-F-13122020	SF13002	İngilizcesi • bişe diyim mi	**bayağı**	iyi. bizden de çok iyiydi. hele	*other*
Y-2-F-03122020-2	SF11004	piyano çalıyodu o. ve şu an gayet iyi.	**bayağı**	iyi çalıyo. yani bilmiyorum.	*other*
Y-3-F-06122020	SF12012		**bayağı**	iyi!	*other*
Y-2-F-13122020	SF13001	edebiyata kıl payı falan giremedi.	**bayağı**	iyidi sıralaması.	*other*
Y-2-F-05122020-1	SF12009	iyi olur. matematiği fullesen zaten	**bayağı**	iyisin sıralama yaparsın. beş on	*interlocutor*
Y-2-F-03122020-2	SF11004	onlar bu arada gerçek hayatta da	**bayağı**	iyi anlaşıyo. hani...	*other*
Y-2-F-05122020-1	SF12009	Endemik'ten çözdüm. şu an geometrim	**bayağı**	iyi.	*self*
Y-2-M-30112020-1	SM09003	model tankları falan var adamların.	**bayağı**	iyi. işte şey • bi de Japonlar hiç	*other*
Y-2-M-05122020	SM11009	sen attmıstın geçen. o	**bayağı**	iyi lan aslında hani.	*other*

L. VITA

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Efeoğlu-Özcan, Esranur

Date and Place of Birth: xxx

Current Affiliation: Gazi University & Council of Higher Education, Turkey

Research Group: Discourse and Corpus Research Group, METU

Email: xxx

ORCID: 0000-0001-9627-4628

OSF: https://osf.io/dgva7/

EDUCATION

2016-2022 Doctor of Philosophy, Language Studies (ELT)
Middle East Technical University, Ankara, Turkey
Dissertation Title: *The Corpus of Turkish Youth Language (CoTY): The compilation and interactional dynamics of a spoken corpus*
Supervisor: Assoc. Prof. Dr. Hale Işık-Güler

2012 - 2015 Master of Arts, English Language Teaching (GPA: 3.88 / 4.00)
Middle East Technical University, Ankara, Turkey
Thesis Title : *The metaphorical reconstruction of Turkey in political discourse: A corpus-driven critical metaphor analysis*
Supervisor: Assoc. Prof. Dr. Hale Işık-Güler

2009 - 2011 Bachelor of Arts, Foreign Language Education (GPA: 3.77 / 4.00)
Minor degree, Psychology
Middle East Technical University, Ankara, Turkey

EMPLOYMENT RECORD

2018 – Present Academic Expert
The Council of Higher Education, Ankara, Turkey

2018 – Present Instructor
Gazi University, Ankara, Turkey

PUBLICATIONS (selected)

Peer-Reviewed Articles

Efeoglu-Özcan, E., (2022). Pull the weeds out or perish: Using pandemic metaphors to strengthen in-group solidarity in Turkish political discourse, *Metaphor and Symbol, 37(*2), 171-84.

Efeoglu-Özcan, E., Erdoğan-Öztürk, Y., Işık-Güler, H. (2021). Why do adults use baby talk in the online space? Baby talk as a pragmatic face device in adult communication, *Çankaya University Journal of Humanities and Social Sciences, 15*(2), 311-329.

Efeoglu, E. & Isık-Güler, H. (2017). Turkey as BODY POLITIC: A comparative perspective on body-related metaphors from Turkish, British, and American political news discourse, *Journal of Linguistic Research, 28*(1), 55-77.

Book Chapters

Baş, M. & **Efeoglu-Özcan, E.** (2022). How to pass this exam? Dealing with COVID-19 through metaphors in Turkish online public discourse, In A. Musolff, R. Breeze, K. Kondo, S. Vilar-Lluch (Eds.), *Pandemic and Crisis Discourse: Communicating COVID-19 and Public Health Strategy* (pp. 207-22). Bloomsbury.

Efeoglu-Özcan, E. (2021). Pragmatics and metaphors. In Ş. S. Çimen & B. Bal-Gezegin (Eds.), *Pragmatics and Language Teaching* (pp. 49-68). Pegem.

PROJECTS (worked as a volunteer, collaborator, researcher)

2020-2022 **Discursive Construction of 'Coronaphobia' in Turkish Media**
GRANT: Scientific and Technological Research Council of Turkey (TUBITAK) 1002 – Scientific and Technological Research Projects Funding Program.
PI: Dr Emel Kökpınar-Kaya (Hacettepe University), **Collaborators**: Dr Yeliz Demir Van Scheppingen (Hacettepe University), Dr Emre Yağlı (Hacettepe University), Dr Oktay Çınar (Istanbul Medeniyet University), Esranur Efeoğlu-Özcan

2018-2021 **Call Center Discourse**
PI: Assoc. Prof. Dr Hale Işık-Güler (METU) **Researchers**: Merve Bozbıyık (METU & Hacettepe University), Esranur Efeoğlu-Özcan

2018-2022 **Türkçe Söz Varlığı Projesi**
Project Lead: The Ministry of National Education/Turkey,
Academic Coordinators for Ankara: Assoc. Prof. Dr Hale Işık-Güler & Esranur Efeoğlu-Özcan

2018-2019 **Women's Coffee Talk**
GRANT: METU Scientific Research Project Fund: GAP-503-2018-3053
PI: Assoc. Prof. Dr Betil Eröz (METU), **Researchers**: Esranur Efeoğlu-Özcan, Banu Çiçek Başaran-Uysal (Eskişehir Osmangazi University)

2008-2011 **The Spoken Turkish Corpus**
Project Lead: Prof. Dr. Şükriye Ruhi, funded by TÜBİTAK & METU Research Fund.
As an undergraduate **project volunteer**, I contributed to the project as a data collector and a transcriber/pragmatic annotator.

ACADEMIC HONORS & AWARDS

2022 METU Publication Award (for SSCI indexed journal article publication)

2016-2022 Graduate Student Scholarship (PhD), Scientific and Technological Research Council of Turkey (TÜBİTAK), Turkey

2012-2015 Dean's High Honor List (MA), Middle East Technical University, Turkey

2012-2015 Graduate Student Scholarship (MA), Scientific and Technological Research Council of Turkey (TÜBİTAK), Turkey

2011-2012 Global Leaders Development Program, Cabinet Office, Government of Japan

2011-2012 Comenius Language Assistantship (Institution: Mankkaanpuron koulu, Espoo, Finland), The European Union

2007-2011 Dean's High Honor List (BA), Middle East Technical University, Turkey

2007-2011 Merit-based Undergraduate Studies Scholarship, Ministry of National Education, Turkey

LANGUAGES

Turkish (C2, Native), English (C2, Fluent), Japanese (B1, Intermediate), French (A2, Upper Beginner)

SOFTWARE

EXMARaLDA: Partitur-Editor, COMA, EXAKT; AntConc; MAXQDA; SPSS

M. TURKISH SUMMARY / TÜRKÇE ÖZET

TÜRKÇE GENÇLİK DİLİ DERLEMİ (COTY): DERLEM OLUŞTURMA VE SÖZLÜ BİR DERLEMİN ETKİLEŞİMSEL DİNAMİKLERİ

GİRİŞ

Zaman içinde veya belirli bir zamanda gözlemlenen dilsel çeşitliliği ve bununla ilişkili olarak da sosyal anlamı araştıran çalışmaların, konuşucu *yaşı*nı bir topluluktaki farklı konuşma gruplarının sınırlarını tasvir etmek ve daha büyük topluluklarına ilişkin ortak ve farklı dilsel pratiklerini keşfetmek için bir parametre olarak kullandığını görmekteyiz. Bu bağlamda gençlerin dilsel pratikleri -*gençlik dili*- dinamik, akışkan ve performatif doğası nedeniyle araştırmacıların ilgisini çekmiştir. Gençlik dili üzerine yapılan dilbilimsel araştırmalarla yalnızca bir dilin güncel kullanımı değil, aynı zamanda dildeki değişimler de araştırılabilmiştir.

'Gençlik' olgusu, sosyal olarak inşa edilmiş bir kategori olduğu için gençlik dilinin tek bir tanımını yapmak mümkün değildir. Bu kavram tarih boyunca farklı kültürel ve siyasi ortamlarda değişen anlamlar ifade etmiştir. Bununla benzer doğrultuda, gençlik dili pratikleri de hiçbir şekilde homojen değildir (Martínez, 2011) ve daha ziyade konuşmacılar tarafından çevrimiçi ve çevrimdışı etkileşimde manipüle edilen dilsel kalıpların ve sosyo-edimsel stratejilerin iç içe geçmiş yönlerini kapsar. Bu dilsel kalıplar ve stratejiler bugüne kadar çeşitli veri kaynakları ve metodolojik yaklaşımlar aracılığıyla araştırılmıştır. Öne çıkan yaklaşımlar arasında toplumdilbilim çalışmalarının yeri oldukça büyüktür. Bu çalışmalar arasında öncül toplumdilbilim araştırmaları bulgularını statik sosyo-demografik kategoriler temelinde genelleştirmişken; son dönem toplumdilbilim çalışmaları doğal ortamında üretilen verileri kullanmakta ve dilsel pratikleri daha akışkan ve performatif bir açıdan değerlendirmiştir. Bu kapsamda dile *söylem* bağlamında yaklaşarak, dil pratiklerini farklı sosyal ve durumsal bağlamlar arasındaki karşılıklı ilişki içinde inceleyen çalışmalar; sözlü söylemin hızlı değişen doğası, gömülü edimsel işlevleri ve parçalı yapısına dikkat çekmektedir (Cutting, 2011).

Dilbilimdeki bu performatif ve söylemsel bakış açısı, konuşucu kategorizasyonlarının ve söylemsel anlam yaratma pratiklerinin akışkanlığını benimserken, dilbilimsel araştırmalarda hedeflenen sistematiklik, şeffaflık, tekrarlanabilirliğin ne şekilde sağlanabileceğine ilişkin endişeleri de beraberinde getirmiştir. Bu noktada *açık bilim* girişimi; dil çalışmalarının geleceği için umut verici bir yol haritası sunmaktadır. Açık bilim şemsiye terimi, bilimsel bilginin -uygun olduğunda- erişilebilir, titiz, tekrarlanabilir, çoğaltılabilir, birikimli, kapsayıcı olması gerektiği fikrine atıfta bulunur (Abele-Brehm vd., 2019; Kathawalla vd., 2020; Syed, 2019; Woelfe vd., 2011). Bu doğrultuda girişim, bilgi yaratma ve yayma konusunda şeffaf ve işbirliğine dayalı yaklaşımların uygulanması çağrısında bulunmaktadır (Fecher & Friesike 2014). *Derlem dilbilimi* de, bu çağrıya bir cevap olarak çok katmanlı dilbilimsel sorgular yürütmek için sürdürülebilir bir araç ve bir dilin sistematik olarak sorgulanması için daha sağlam bir sistem sunmaktadır. Dil kullanımına ilişkin geniş kanıtlar bütününü oluşturan *derlem*ler sayesinde (İng. corpus) (McEnery, 2005, 2012), dilsel örüntülerin durumsal ve sosyal değişkenlerle ilişkili olarak incelenmesine yönelik tutarlı ve güvenilir bir araç ortaya çıkmıştır. Bu doğrultuda, bu çalışma derlem dilbilimini Türk gençleri tarafından konuşulan çağdaş Türkçenin incelenmesi için kullanarak dilbilimde tutarlı, sürdürülebilir, hesap verebilir araştırmalara katkıda bulunmak için açık bilim uygulamalarına bağlı kalmaktadır. Bu çalışma, Türkçe için oluşturulan ilk gençlik dili derlemi olan Türkçe Gençlik Dili Derlemi'nin (CoTY) derlenmesi ve özgün dil verilerinin derlem dilbilim araçlarını kullanarak sistematik olarak incelenmesi yoluyla, gençlik etkileşiminin çoklu etkileşimsel yönlerini incelemek için bir temel ve öncül bir çalışma sunmaktadır.

ALANYAZIN

Gençlerin dil pratiklerine yönelik ilk toplumdilbilimsel çalışmalar, gençlerin dilindeki düzenliliklerin rutinleşmiş ve sistematik tanımını belgelemek için Labov'un yerel dil kavramını (İng. vernacular) takip etmekle beraber bu yaklaşım gençlik dilinin ait oldukları toplumdaki ana akım standart dille karşılaştırılmasına sebep olmuş ve gençlik dilinin eksik, tamamlanmamış veya bir geçici dil pratikleri bütünü olarak ele alınmasına yol açmıştır. Bu tartışmalara bir yanıt olarak Kotsinas (1998) gençlerin dilinin karmaşıklığını ve heterojenliğini ele almak için Stockholm gençliğinin dilsel pratiklerini şehirdeki diğer çeşitlerle birlikte ayrı bir çeşitlilik olarak tasvir etmek için *multiethnolect*

terimini ortaya atmıştır. İskandinav araştırma bölgelerinde gelişen bu çeşitlilik yaklaşımı, multiethnolect'in göçmen gençler tarafından kullanıldığını ve çeşitli miras dillerden gelen bir dizi dilsel biçim ve uygulamanın yerel toplumun ana akım diliyle karıştırılmasıyla karakterize edildiğini öne sürmüştür. Ancak yerel dillerden farklı olarak, çoklu diller kasıtlı ve belirgin olabilir (Cheshire vd., 2015). Çokdillilikle ilgili olarak Cheshire, Kerswill, Fox ve Torgersen (2011), Londra'nın iç kesimlerindeki çoklu etnik kimliklere sahip genç konuşma toplulukları tarafından gerçekleştirilen dilsel çeşitliliğe odaklanmış ve konuşmacıların kullandığı ayırt edici diller arası özelliklerin repertuarına atıfta bulunmak için *Multilingual London English* terimini ortaya atmıştır. Bununla birlikte, multiethnolect terimi, dilin edimselliğini göz ardı ettiği ve etnik kökenle ilgili bir konumlandırmayı ima ettiği için sıklıkla eleştirilmiştir. Dorleijin ve Nortier (2015), buna bir yanıt olarak, gençlerin dilsel pratiklerindeki stilizasyonun etkileşimini vurgulamış ve *urban youth speech style* terimini önermiştir. Bu konu, Rampton (1995) tarafından da geniş bir şekilde tartışılmış ve çok etnikli gençlik topluluklarının etkileşimsel pratiklerini bir üslup pratiği olarak ele alan *language crossing* terimini ve daha güncel çalışmalarında ise (Rampton, 2011, 2013, 2015) *contemporary urban vernacular* terimini kullanmıştır. Benimsenecek terminoloji konusunda bir fikir birliği olmasa da, yaklaşımların çeşitliliği, daha büyük bir bütünün tamamlayıcı yönlerini araştırmak için farklı bakış açılarına katkıda bulunmaktadır. Son zamanlarda yapılan çalışmaların, özellikle gençler arasındaki söylemsel etkileşimin bağlama bağımlılığını, çokluğunu ve dinamik özelliklerini vurgulamak için *gençlik dilleri* (İng. youth languages) terimini savunduğunu belirtmek gerekir. Bu görüşte ortak payda hala biyolojik yaştır, ancak yaş spektrumunun sınırları önceden belirlenmiş bir aralığa sahip değildir. Bu kapsayıcı görüş doğrultusunda, bu çalışma da derlem verilerini tanımlamak, keşfetmek ve analiz etmek için gençlik dili ve -özellikle sözlü etkileşim için- gençlik konuşması terimini benimsemiştir.

Alanyazında gençlerin dilsel pratiklerini incelemek için iki yaygın araştırma yaklaşımı bulunmaktadır: varyasyonist (İng. variationist) çalışmalar ve derlem odaklı çalışmalar. Varyasyonist çalışmalar; genç konuşmacıların dilsel pratiklerini sistematik ve yapılandırılmış bir olgu olarak inceler ve konuşmalarındaki çeşitlilik örüntülerini tespit etmeyi amaçlar. İlk dalga varyasyonist çalışmalar sosyo-demografik kısıtlamalar, özellikle de sosyal sınıf, cinsiyet ve konuşmacıların yaşı açısından açıklamayı amaçlamıştır (Trudgill, 1974, 1983; Labov, 1972). İkinci dalga çalışmalar ise doğal olarak ortaya çıkan verilerden ve etnografiler gibi daha niteliksel yönelimli metodolojilerden

faydalanmıştır. Birinci dalga araştırmalardaki sosyal anlamın deterministik görüşünün aksine, ikinci dalga çalışmalar yerel dil kullanımında konuşmacı etkisini ve iç içe geçmiş sosyal kategorileri vurgulamıştır (Cheshire, 1982; Eckert, 1989, 2000). Üçüncü dalga çalışmalara gelindiğinde ise, kimlik ve ideoloji gibi olguların konuşucuların inşa ettiği söylemlerle nasıl dinamik olarak ortaya çıktığına odaklanılmaya başlanmış ve genç konuşucuların dilsel pratikleri bağlamsal ve etkileşimsel boyutlarda incelenmeye başlanmıştır (Androutsopoulos, 2015; Bodén, 2004, 2011; Eckert, 2000; Ilbury, 2019; Jørgensen, 2008; Madsen, 2015; Sierra, 2016).

Varyasyonist çalışmaları tamamlayıcı ve destekleyici nitelikte olan derlem çalışmaları ise, geçtiğimiz son otuz yıl içinde giderek artmıştır. Bu çalışmalar varyasyonist çalışmalarla örtüşen araştırma gündemlerine sahip olsalar da, gençlik dili araştırmaları içinde sürdürülebilirlikleri, diller arası karşılaştırılabilirlikleri ve sundukları topluluğun temsili bir örneğini sunmaları açısından öne çıkmaktadırlar. Derlem dilbilimi ve gençlik dili çalışmalarının kesiştiği noktada yürütülen çalışmaların özellikle İngilizce, İspanyolca, Almanca, Danca, Fince, İzlandaca, Norveççe ve İsveççe üzerinde yoğunlaştığı görülmektedir. Gençlik dili derlemlerinin ilk örneği, Stenström ve ekibi (Stenström vd., 2002) tarafından 1993 yılında oluşturulan 444,166 sözcüklük *The Bergen Corpus of London Teenage Language (COLT)* derlemidir. İngiliz Ulusal Derlemi'nin (BNC1994) bir alt derlemi olan bu derlem, Londra'da yaşayan ve İngiliz İngilizcesi konuşan 13-17 yaş aralığındaki 33 gencin sözlü verisinden oluşturulmuştur. Bu derlemin oluşturulmasıyla genç konuşucular tarafından konuşulan İngilizcede gözlemlenen çok çeşitli dilsel araçlar (söylem belirteçleri, küfür ve argo ifadeleri, pekiştireçler, vb.) ve olgular (çatışma etkileşimi, hikayeleştirme, vb.) üzerinde dilbilimsel analizler yapılması sağlanmıştır. Konuşucular arasında yalnızca gençlerin olmaması ve aile, öğretmen etkileşimine yönelik verinin de dahil edilmesi, tüm verinin etkileşimsel değil aynı zamanda monolog verileri de içermesi, ve üstverinin (İng. metadata) tüm konuşucular için işaretlenmemiş veya eksik işaretlenmiş olması gibi sınırlılıkları olmasına rağmen, COLT sonraki gençlik dili derlemleri için bir temel örnek teşkil etmektedir. COLT'la ilişkilendirilebilecek derlem çalışmaları arasında en öne çıkanı İspanyol gençlik dili için hazırlanmış olan *Corpus Oral de Lenguaje Adolescente (COLA)* derlemidir. Bu derlem Madrid'de yaşayan 13-18 yaş aralığındaki 145 gencin sözlü verisi ile oluşturulmuş ve derlem hem İspanyolca dilinin güncel kullanımına hem de İspanyol ve İngiliz gençlik dillerinin karşılaştırmalı dilbilimsel analizlerinin yapılmasına imkan vermiştir (Drange, 2009; Stenström, 2007, 2014; Stenström & Jørgensen, 2009; Jørgensen, 2008, 2009, 2013).

Yakın zamanda derlenen *Corpus Oral de Madrid (CORMA)* derlemi ile ise COLA ile karşılaştırmalar yapılarak İspanyolca gençlik dilindeki zamana bağlı dilsel değişikliklerin saptanması mümkün kılınmıştır (Enghels et al., 2020; Roels, 2021). Almanca gençlik dili derlemleri arasında yer alan *Ph@ttSessionz Projesi* (Draxler et al., 2008) 12-20 yaş aralığındaki Almanca konuşan gençlerin kullandığı dilin fonetik özelliklerini saptamayı amaçlamış; *KiezDeutsch-Korpus (KiDKo)* 14 – 17 yaş aralığındaki farklı etnik kimliklere sahip genç Almaca konuşucularının kod-değiştirme (İng. code-switching) pratiklerini, dil kullanımlarına dair tavır ve ideolojilerini incelemek istemiş (Rehbein et al., 2014); 14 -1 7 yaş aralığındaki İsviçre Almancası konuşan gençlerin sözlü verisiyle oluşturulan *Jugendsprache Schweiz Korpus (JuBE)* ise gençler arasındaki çokdillilik pratiklerine odaklanmıştır (Schneider et al., 2021). Son olarak İskandinav gençlik derlemleri, bu ülkelerde yaşanan sosyopolitik gelişmelerle benzer şekilde göç ve dil etkileşimleri odağında ortaya çıkmıştır. Örneğin *UNO* derlem projesi (*Språkkontakt och ungdomsspråk i Norden*), Danimarka, Finlandiya, İzlanda, Norveç ve İsveç'te yaşayan gençlerden toplanan veri ile argo kullanımındaki güncel örüntüler ve göçmen dillerinin bu örüntüler içindeki görünümleri çalışılmıştır (Drange, 2002). Tüm bu derlemlere ek olarak, yine genç konuşuculardan elde edilen verilerle oluşturulan ama pedagojik amaçlar taşıyan öğrenci derlemleri (İng. learner corpora) de mevcuttur. Bu derlemler arasında en öne çıkan projelerden biri olan *International Corpus of Learner English (ICLE),* 5.5 milyon sözcükten oluşan ve 25 farklı anadile sahip İngilizce dili öğrencisinin verisiyle oluşturulmuş bir derlemdir. Bu derlemin sözlü derlem bileşeni olarak *Louvain International Database of Spoken English (LINDSEI)* ise ileri seviyede İngilizce dil yetkinliğine sahip lisans öğrencilerinin sözlü verisinden oluşmaktadır. Benzer amaçla oluşturulan *The System Aided Compilation and Open Distribution of European Youth Language (SACODEYL)* de dil öğrenimi ve öğretimi için pedagojik bir kaynak sunmak ve dil edinimine yönelik veri odaklı yaklaşımları kolaylaştırmak amacıyla oluşturulmuş bir derlemdir (Pérez-Paredes & Alcaraz-Calero, 2009).

Türkiye'deki gençlik çalışmalarına döndüğümüzde, bu alanın yeni bir araştırma alanı olmadığını, bununla beraber dilbilimsel çalışmaların ne yazık ki oldukça sınırlı olduğunu görmekteyiz. Türkiye'de gençlik üzerine yapılan araştırmaların büyük bir kısmı eğitim, psikoloji ve sosyoloji alanlarında gerçekleştirilmiştir. Sosyoloji çerçevesinde bakıldığında Demir (2012, s. 98), Türkiye'de 1980-2000 yılları arasındaki gençlik çalışmalarının çoğunlukla bağlamsal farklılıkları ve disiplinler arası modelleri göz ardı ettiğini, çoğunlukla anketlere dayandığını ve kurumsal ya da devlet fonlarından ziyade

bireysel araştırmacılar tarafından yürütüldüğünü belirtmektedir. 2000 yılından itibaren ise araştırmaların daha disiplinler arası hale gelmesiyle, Türkiye'deki gençlik çalışmaları pek çok farklı olguyu incelemiştir. Bu güncel odaklar arasında gençlerin değişen sosyal ve kültürel alışkanlıkları (örn. Özensel, 2009; Yazıcı, 2001), gençlerin çevrimiçi kanallar aracılığıyla katılımı (örn. Neyzi, 2011; Telli-Aydemir, 2009), çevrimiçi gençlik kültürleri (örn. Tuzcu Tığlı, 2019), ortaya çıkan kimlikleri ve toplumsal cinsiyeti (örn, Alemdaroğlu, 2007; 2010; Çelik ve Lüküslü, 2010; Demez, 2009; Yonucu, 2005), genç işsizliği (örneğin, Yentürk ve Başlevent, 2008; Yücel ve Lüküslü, 2013), gençlik alt kültürleri (örneğin, Semerci, Erdoğan ve Sandal Önal, 2017; Şişman, 2013) ve son zamanlarda iklim aktivizmi ve gençlik (örneğin, Atik ve Doğan, 2019) sayılabilir.

Dilbilim ve dil öğretimi alanlarındaki çalışmaları incelendiğinde ise çalışmaların büyük kısmının varyasnonist paradigmada konumlandığı ve ilk çalışmaların oldukça sınırlı veriye dayandırıldığı görülmüştür. İlk dilbilimsel gençlik dil pratikleri incelemesi olan Açıkalın'ın (1991) çalışması, 17-19 yaş arası Türkçe konuşanların evde konuştukları dil ile akran grupları arasında konuşurken kullandıkları dilin farklılıklarına odaklanmış ve her kuşağın farklı bir dil kullanımına sahip olduğunu savunmuştur. Katılımcıların profili ve toplanan verilerin kapsamı hakkında sınırlı bilgi olsa da, çalışmanın dikkat çekici yanı, doğal olarak ortaya çıkan veriyi kullanmasıdır. Bu çalışma dışındaki çalışmaların büyük çoğunluğu araştırmacı tarafından elde edilmiş (İng. elicited data) ve kapalı uçlu anketler ve yapılandırılmış görüşmeler gibi oldukça sınırlandırıcı veri toplama metotları kullanmıştır (örneğin, Şafak ve Bilginsoy, 2019; Toğrol, 2012). Bu çalışmaların ortak noktası gençlik dilini standart dil dışı bir kullanım, eksik ve yanlış dil pratikleri olarak gören kuralcı (İng. prescriptivist) bir bakış açısı sunmalarıdır. Türkçe üzerinde yapılan çalışmalar arasında gençlik dili her ne kadar kimi zaman gelip geçici bir stil (Gunay, 2007) veya Türkçe'yi bozan dil kullanımları bütünü (Canbulat, 2017; Kırık, 2012) olarak gören bakış açıları bulunsa da, bu görüşleri destekleyici nitelikte dilbilimsel kanıta dayalı bilimsel bir çalışma henüz bulunmamaktadır.

YÖNTEM

Türkçe gençlik dilinin ilk derlemini oluşturma amacını güden bu çalışma, Çağdaş konuşma Türkçesinde ikili ve çok partili gençlik etkileşiminin dilsel özelliklerini ve söylemsel dinamiklerini incelemek için bir sözlü özel alan derlemi oluşturmuştur. Bu çalışmanın araştırma soruları iki ana katmanda toplanmaktadır. Birinci katman *derlem*

*oluşturma*ya ilişkindir ve CoTY'nin yapısal bileşeninde yer alan çeşit (İng. type) ve örnekçe (İng. token) sayısı, konuşucuların demografik dağılımı, verinin konuşmacı ve konuşmacı gruplarına göre dağılımını tasvir etmeyi amaçlar. İkinci katman, verinin *dilsel özellikleri*ni makro ve mikro boyutta ele alarak, makro boyutta derlemde gözlemlenen ana konuları, alt konuları, anahtar kavramları (İng. keywords); mikro boyutta ise dört grup etkileşim belirleyicisinin (*yansıma belirteçleri*, *hitap sözcükleri*, *belirsizlik ifadeleri*, *pekiştiriciler*) türlerini, derlemdeki sıklık dağılımlarını, işlevlerini araştırmaktadır.

Derlemin ait olduğu evreni azami azami düzeyde temsil edici bir örneklemi olabilmesi için çeşitli yaklaşımlar gerçekleştirilmiştir. Öncelikle, toplanacak dilsel verilerin sınırları net bir şekilde tanımlanmış, yalnızca konuşma diline ait sözlü veri toplanmıştır. İletişim biçimi spontane, arkadaş arasında, yüz yüze veya çevrimiçi etkileşim olarak nitelendirilmiştir. Konuşmalar hem iç mekan (örn. yatak odası, mutfak, oturma odası) hem de dış mekanları kapsayan (örn. sokak, park, evin arka bahçesi) gayri resmi ortamlarla sınırlıdır. Derlemdeki konuşuların tamamı, 14-18 yaş aralığındaki ana dili Türkçe olan lise öğrencilerinden oluşmaktadır. Veri toplama süreci iki aşamada gerçekleşmiştir. Öncelikle ilk aşamada örneklem kriterlerine uygun gençler, kolaylıkla bulunabileni örnekleme yöntemi ve devamındaki kartopu örnekleme ile (İng. convenience sampling ve snowball sampling) çalışmaya davet edilmiş ve bu katılımcılardan elde edilen veri 'ilk parti' kayıtlarını oluşturmuştur. İlk parti kayıtlar, araştırmacının örneklemdeki dengesizlikleri tespit etmesi ve ilk partiyi tamamlayıcı nitelikteki katılımcılardan oluşan 'ikinci partiyi' düzenlemesi için referans noktası olarak işlev görmüştür, böylece maksimum varyasyon örneklemesi (İng. maximum variation sampling) kullanılarak konuşmacıların cinsiyeti ve her sınıf seviyesindeki katılımcı sayısı arasındaki denge çeşitlendirilmiştir.

Araştırmacı, katılımcılarla yüz yüze veya çevrimiçi kanallar aracılığıyla görüşerek çalışma kapsamını açıklamış, veri toplama sürecine ilişkin onları bilgilendirmiştir. Katılımcıların bu derlem çalışmasında dört ana rolü olmuştur: (i) akranlarıyla yaptıkları konuşmayı kaydederek araştırmacıya sunmak, (ii) bir Kayıt Günlüğü doldurarak konuşmadaki tüm konuşmacılar hakkında demografik bilgi ve ayrıntılı üstveri sağlamak, (iii) yeni potansiyel katılımcılara çalışmayı tanıtmak, (iv) konuşmada karşılaşılan muğlak veya anlaşılmaz ifadeleri netleştirmek için araştırmacıya emik bir bakış açısı sağlamak ve gerektiğinde yorumların doğrulanması için araştırmacıyla iletişim halinde olmak. Katılımcılar iletişimsel veriyi telefonları veya bilgisayarlarının ses kayıt özelliğini

kullanarak kaydetmiş, ardından bu veriyi Kayıt Günlüğü ile hem katılımcılar hem de ebeveyn/vasilerden alınan onam formları ile beraber araştırmacıya iletmişlerdir.

Sözlü verinin çeviriyazıya aktarılması, üstverinin işlenmesi, işaretlemelerin (İng. annotation) yapılması, çeviriyazının zamana göre hizalanması için bir derlem oluşturma programı olan EXMARaLDA (Schmidt & Wörner, 2014) kullanılmıştır. Bu program, Partitur-Editor ve COMA araçlarından yararlanılarak derlemin oluşturulması aşamasında kullanıldığı gibi EXAKT aracının kullanılması ile de derlemde sorguların ve dilsel analizlerin yapılabilmesi sağlanmıştır. Çeviriyazı yönergesi için HIAT (Rehbein vd., 2004) ve STD (Ruhi, Hatipoğlu, Işık-Güler, & Eröz-Tuğa, 2010) çeviriyazı rehberleri bu çalışmanın verisine göre adapte edilmiştir.

Bu çalışmanın ikinci aşamasını oluşturan gençlik dilindeki etkileşimsel dinamiklerin incelenmesi ise derlem odaklı söylem çalışmaları (İng. Corpus-oriented discourse studies) eksenine oturtulmuştur (Gabrielatos, 2021). Bu çalışmanın odağında yer alan dört etkileşim belirleyicisi olan (i) *yansıma belirteçleri* (İng. response tokens), (ii) *hitap sözcükleri* (İng. vocatives), (iii) *belirsizlik ifadeleri* (İng. vague expressions) ve (iv) *pekiştiriciler* (İng. intensifiers), sıklık analizi, bağlam içinde anahtar sözcük (İng. KWIC), eşdizim (İng. collocation) gibi derlem analiz teknikleri kullanılarak incelenmiş ve söylem içinde öne çıkan edimsel görevleri incelenmiştir.

BULGULAR VE TARTIŞMA

Bu çalışma, birbirini tamamlayan iki amaç çerçevesinde yürütülmüştür. Birinci amaç, genç yaştaki Türkçe konuşucularının dilsel pratiklerini incelemek için sürdürülebilir bir araç -bir derlem- oluşturmak, ikinci amaç ise bu aracı kullanarak bu konuşmacılar arasındaki sözlü etkileşimin göze çarpan özelliklerini araştırmaktır. Bu amaçlar doğrultusunda, bulgular iki katman altında özetlenecektir: derlem oluşturma ve etkileşimsel dil pratikleri.

Derlem yapısına ilişkin bulgular

Oluşturulan derlem, CoTY, Ekim2019 ile Ekim 2021 arasındaki süre zarfında maksimum varyasyon örnekleme kullanılarak toplanan 14-18 yaş aralığında anadili Türkçe olan liseli gençlerin arkadaşları arasındaki spontane ve doğal sohbetlerini içeren 26 saat ve

11 dakikalık sözlü veriden oluşmaktadır. Derlemdeki konuşucular iki veya üç kişilik gruplarda, yüzyüze veya çevrimiçi iletişim araçlarını kullanarak iletişim kurmuştur. Örneklemde 25 şehirden, toplam 123 konuşucu (62 kadın ve 61 erkek) konuşucu yer almaktadır. Derlemde 168748 örnekçe (İng. token) ve 24736 çeşit (İng. type) sözcük bulunmaktadır.

Derlem konuşucu cinsiyeti açısından dengeli bir dağılım gösterse de, sözlü verinin dağılımında gözlenen çarpıklık (İng. skewed) verinin araştırmacı müdahalesi olmadan toplandığına ve özgünlüğüne işaret olarak görülmektedir. Bu kapsamda verinin dağılımına bakıldığında, kadın konuşucuların 97676 örnekçe (derlemin 58%'sine karşılık gelmektedir), erkek konuşucuların ise 71072 örnekçe ürettiği görülmüştür. Derlemde üç çeşit konuşmacı grubu bulunmaktadır; yalnızca kadınların bulunduğu gruplar, yalnızca erkeklerin bulunduğu gruplar, hem kadın hem erkek konuşmacıların bulunduğu gruplar. Bu konuşmacı türlerine göre very dağılımına bakıldığında; yalnızca kadınların bulunduğu konuşmacı gruplarında 84076 örnekçe, yalnızca erkeklerin bulunduğu konuşmacı gruplarında 43849 örnekçe, karışık konuşmacı gruplarında ise 40823 örnekçe olduğu görülmüştür.

Derlemin makro ve mikro özelliklerine ilişkin bulgular

Daha önce de bahsedildiği gibi, derlemin dilsel özellikleri makro ve mikro boyutta incelenmiştir. Makro boyutta yapılan inceleme, derlemdeki ana konuları, alt konuları ve anahtar kavramları ortaya koymuştur. Bu kapsamda, CoTY'de 11 ana başlık altında kümelenmiş toplam 47 konuşma konusu tespit edilmiştir. Sonuçlar, konuşmacıların izledikleri programlar, okul ödevleri, günlük işler gibi günlük konuların yanı sıra romantik ilişkiler, politika, aile sorunları, akıl sağlığı ve cinsellik gibi özel ve hassas konuları da içeren geniş bir yelpazeyi kapsayan bir dizi konu hakkında konuştuklarını göstermiştir. En sık konuşulan konular, *eğlence* ana konusu (%23) etrafında kümelenmektedir. Bu konunun altında, konuşucuların seyrettikleri filmler, diziler, programlar; takip ettikleri sosyal media programları, hesapları, figürleri; okudukları kitaplar ve beğendikleri yazarlar; dinledikleri sanatçılar, müzik tarzları yer almaktadır. En sık bahsedilen ikinci konu, genç konuşmacıların çeşitli etkileşimsel belirteçleri de kullanarak muhataplarıyla duygusal bağ kurmalarını sağlayan *sosyal ve duygusal bağlar* konu başlığıdır (%20). Bu konu kümesi altında konuşucular, duygusal bağ kurdukları arkadaşları, öğretmenleri, aileleri gibi yakın çevrelerindeki kişiler hakkında olduğu

kadar hayranlık duydukları fakat tanımadıkları kişiler hakkında da konuşmaktadır. Üçüncü en büyük konu kümesi ise *eğitim*dir (%17) ve bu konu başlığı altında ağırlıklı olarak konuşmacıların eğitim sisteminde karşılaştıkları sorunlar, çalışma rutinleri ve akademik hedef ve hayalleri yer almaktadır. Tüm bu ana konular, konuşmacılar etkileşimi ortaklaşa inşa ederken onlara ortak bir kavramsal alan ve dilsel ve semiyotik kaynaklar repertuarı sağlamaktadır.

Derlemin sözcüksel özellikleri, derlem için sözcük listesi (İng. wordlist) oluşturularak ve bu belirteçlerin sıklıkları referans derlem STC'deki sıklıklarıyla karşılaştırılarak belirlenmiştir. Bu şekilde anahtar kelime analizi (İng. keyness analysis) kullanılarak, Türk gençlik konuşmaları için tipik olan olumlu anahtar kelimeler belirlenmiştir. Sonuçlar iki grup anahtar kelime ortaya çıkarmıştır. İlk grupta, *günlük yaşam* ve *eğitim* kavramsal alanları ile ilişkilendirilen sözcüklere atıfta bulunan anahtar kavramlar yer almaktadır. İkinci gruptaki anahtar sözcükler ise bu çalışma kapsamında *etkileşim belirleyicisi* (İng. interactional marker) olarak etiketlenen (Ruhi, 2013) işlev sözcüklerini kapsamaktadır. Bu gruptaki dilsel öğeler, söylemde sosyo-edimsel işlevler sergilemekte olup daha önce de belirtildiği gibi dört kategoriye ayrılmıştır: *yansıma belirteçleri*, *hitap sözcükleri*, *belirsizlik ifadeleri*, *pekiştiriciler*. Bu çalışmadaki tüm etkileşimsel beliryecileri EXMARaLDA'nın EXAKT derlem aracı kullanılarak çağrılmış; sıklık analizi, bağlam içinde anahtar sözcük, eşdizim derlem teknikleri kullanılarak analiz edilmiştir. Her bir etkileşim belirleyici kategorisi derlemde ayrı ayrı incelenmiş, derlemdeki dağılımları ve öne çıkan edimsel işlevleri tartışılmıştır. Aşağıda her kategoride öne çıkan bulgular sunulmaktadır.

Derlemdeki Etkileşim Belirleyicileri: Yansıma Belirteçleri

Yansıma belirteçleri (İng. response tokens) iletişimde aktif dinleyiciliği gösteren yanıt belirteçleridir. İngilizce için yapılan sınıflandırmalarda çoğunlukla minimal ve minimal olmayan ayrımı kullanılırken (Fellegy, 1995; Fishman, 1978; Gardner, 1997, 2001; Schegloff, 1982; Tottie, 1991), bu çalışmada Türkçenin hem morfolojik hem de edimsel özelliklerine dayanan farklı bir sınıflandırma önerilmiştir. Bu sınıflandırma, kısa seslendirmeleri ve ünlemleri içeren *sözlüksel olmayan yansıma belirteçleri* (İng. non-lexical response tokens) ile tek kelimelik sözlüksel tepkileri ve bu tepkilerin tekrarları gibi kısa sözlüksel tepki kümelerini içeren *sözlüksel yansıma belirteçleri*nden (İng. lexical response tokens) oluşmaktadır.

CoTY'de 36 türde toplam 1305 sözlüksel olmayan yansıma belirteci ve 37 türde toplam 1728 sözlüksel yansıma belirteci tespit edilmiştir. Derlemde en sık gözlenen sözlüksel olmayan yansıma belirteci *hı-hı* (*AF*=337, *RF*=1997.06), en sık gözlenen sözlüksel yanıt belirteci ise *evet* (*AF*=1582, *RF*=9374.93) olarak tespit edilmiştir. Geleneksel yansıma belirteçlerine ek olarak, tabu dil alanlarından sözcüksel öğeler, dini terminolojiye ait sözcükler, ve görece güncel argo tabirler de Türkçe konuşan gençler tarafından yansıma belirteci olarak kullanılmıştır. Türkçe gençlik konuşmalarında yansıma belirteçlerinin edimbilimsel işlevlerini araştırmak amacıyla, derlemde en sık rastlanan ikinci sözcüksel yansıma belirteci *aynen* (*AF*=329, *RF*= 1949,65) analiz odağı olarak seçilerek daha detaylı incelenmiş ve derlem bağlamı içinde kullanımları tartışılmıştır. Derlem analizi, *aynen* sözcüğünün işlevinin Türkçe gençlik konuşmalarında geleneksel olarak öngörülen zarf kullanımının ötesine geçtiğini göstermiştir. Bu sözcük, CoTY'de belirgin bir şekilde yansıma belirteci olarak işlev görmektedir. O'Keefe ve Adolphs'un (2008) taksonomisine dayanan analizle *aynen*'ın en sık olarak söylemin akışını sürdürmek ve mevcut konuşmacıyı konuşmaya devam etmeye teşvik etmek için bir *devam ettirici* (İng. continuer) (%47) olarak kullanıldığını göstermiştir. İkinci en sık görülen işlev, literatür tarafından anlaşma(sızlık)ı ve konu değişimini işaret ettiği bildirilen *yakınsama* (İng. convergence) (%25) işlevidir. *Bağlılık* (İng. engagement) işlevi (%15), dinleyicilerin muhatapları tarafından iletilen mesajlara duygusal bağlılık göstermelerini sağlamaktadır. Son olarak, *aynen* belirteçlerinin en küçük oranı (%13), muhatabın daha önceki bir içerik veya mesaj açıklamasını anladığını teyit etmek için kullanılan *bilgi alma belirteçleri* (İng. information receipt) olarak tanımlanmıştır. Genel olarak sonuçlar, *aynen*'ın konuşma Türkçesinde birden fazla edimsel işlevi olduğunu ve özellikle gençlik konuşmalarında belirgin olduğunu kanıtlamıştır. Bu belirginlik, mevcut diğer Türkçe derlemlerde *aynen* için derlem sorguları da yapılarak teyit edilmiştir. Bu kapsamda, 2008-2013 dönemini kapsayan Türkçe yetişkin konuşmaları verisi sağlayan Sözlü Türkçe Derlemi (STD) ve çağdaş yazılı (ve kısmen sözlü) Türkçenin genel bir derlemi olan Türkçe Ulusal Derlemi (TUD)'dir. STD ve TUD verileri ile karşılaştırıldığında, *aynen*'ın CoTY'de daha sık olduğu (göreceli frekanslar CoTY'de *RF*=1949,65, STD'de *RF*=195,53 ve TUD'da *RF*=66,83'tür), dolayısıyla bu sözcüğün Türkçe gençlik konuşması için kayda özgü bir belirteç ve aynı zamanda konuşma Türkçesinde son zamanlarda görülen bir eğilimin işaretçisi olarak değerlendirilebileceği görülmüştür.

Derlemdeki Etkileşim Belirleyicileri: Hitap Sözcükleri

Alanyazında yakın arkadaşlar arasında hitap sözcüklerinin kullanılmadığı vurgulanmış olsa da (Biber vd., 1999), gençlik dili üzerine yapılan son çalışmalar, hitap sözcüklerinin kullanımının genç konuşmacılar arasındaki etkileşimin belirgin bir özelliği olduğunu göstermektedir (Palacios Martínez, 2011, 2021; Parkinson, 2020; Rendle-Short, 2009, 2010; Roels vd., 2021; Stenström vd., 2002). Bu çalışmada özellikle nominal hitap sözcüklerine odaklanılmış, zamir ve kişi adları kapsam dışı bırakılmıştır.

Derlem analizinde, CoTY'de toplam 48 türde 2111 hitap sözcüğü tespit edilmiştir. En sık rastlanan hitap sözcüğü *kanka* ve varyantları *kanki, kank, kanks* (*AF*=680, *RF*=4029.67); ardından *oğlum* (*AF*=452, *RF*=1789.65); ve *abi* (*AF* =302, *RF*=1789.65) olmuştur. Derlem verilerine dayanarak yapılan gözlemler esas alındığında, hitap sözcüklerini (Biber vd., 1999 tarafından önerildiği gibi) 'sevgi' (İng. endearment), 'tanıdıklaştırma' (İng. familiarizer) ve 'hakaret' (İng. insult) gibi orijinal semantik kategoriler açısından sınıflandırmanın, bu belirteçlerin edimsel işlevlerini açıklamak için nispeten dar bir yaklaşım sunduğu vurgulanmıştır. Ayrıca, hakaret sözcüklerinin hem kadın hem de erkek konuşmacılar tarafından kullanıldığı ve bu sözcüklerin hem aşağılama hem de sosyal bağ kurma amacıyla her tür konuşmacı grubu (kadın-kadın, erkek-erkek ve karma) arasındaki etkileşimlerde gözlemlendiği ortaya çıkmıştır. Hitap sözcüğü olarak kullanılan hakaret sözcükleri grubunun, sahip olduğu sözcük türü sayısı bakımından en zengin hitap sözcüğü kategorisi olması da dikkat çekicidir (*n*=14). Hitap sözcüklerinin göndergelerine ilişkin bir başka gözlem de, bazı hitap sözcüklerinin Türkçede anlamsal olarak cinsiyete göre işaretlenmiş olmasına rağmen, konuşmacıların bunları CoTY'de hem kadın hem de erkek muhataplara hitap etmek için kullandıkları bulgusudur. Tüm bu gözlemler, gençlik konuşmalarındaki hitap sözcüklerinin edimsel genişleme (İng. pragmatic extension) gösterdiğini ve dolayısıyla işlevlerinin belirlenmesinin bağlamsal ve ilişkisel bir yaklaşım gerektirdiğini ortaya koymuştur. Bu amaçla, McCarthy ve O'Keeffe'nin (2003) işlevler için organizasyonel ve kişilerarası düzeyler taksonomisi kullanılarak derlem analizinde en sık kullanılan hitap sözcüğü olan *kanka* (*AF*=680, *RF*=4029.67) sözcüğü işlevlerine odaklanılarak detaylı bir şekilde incelenmiştir.

Buglular; konuşmacıların *kanka*'yı kişilerarası amaçlara (*n*=306) kıyasla organizasyonel amaçlar (*n*=374) için nispeten daha fazla kullandığını ortaya koymuştur. Tüm alt işlevler söz konusu olduğunda ise, *kanka*'nın konuşma sırası yönetimi (İng. turn management),

konu yönetimi (İng. topic management), çağrılar (İng. summons) işlevleri ile kişilerarası işlevler altında şakalaşma (İng. badinage), yumuşatma (İng. mitigatory), ilişkisel işlev (İng. relational) olmak üzere tüm alt işlevler için kullanıldığını ortaya koymuştur. Alt işlevler açısından bakıldığında, hitap sözcüğü olarak *kanka* en sık etkileşimde konuyu başlatma, genişletme, değiştirme, kapatmayı kapsayan *konu yönetimi* (%32) amacıyla kullanılırken; bu işlevi, kişisel değerlendirmeleri, anlaşmaları, yüz güçlendiricileri aktarmak için kullanılan *ilişkisel* amaç (%18) ve derlemdeki konuşmacıların olumlu/olumsuz yüzüne (İng. positive/negative face) yöneltilen potansiyel tehditleri hafifleten *yumuşatma* işlevi (%15) izlemiştir.

Derlemdeki Etkileşim Belirleyicileri: Belirsizlik İfadeleri

Konuşmacılar arasında paylaşılan kavramsal alanı (İng. shared conceptual space) yansıtmak için kullanılan belirsizlik ifadeleri (İng. vague expressions) olarak bu ifadeler, daha önce gayri resmi ve samimi konuşma kayıtlarındaki sıklıkla kullanıldıkları belirtilmiştir (Clancy, 2016; Evison vd., 2007; Stenström vd., 2002). Bu çalışmada, belirsizlik ifadeleri iki grup altında incelenmiştir: belirsiz göndermeler (İng. vague references) ve belirsiz eklentiler (İng. vague additives).

Bulgular CoTY'de 26 tür ve 4438 belirsizlik ifadesi tespit edilmiştir. Belirsiz göndermeler, tespit edilen belirsizlik ifadelerinin büyük bir bölümünü (%68) oluşturmaktadır. Bu belirsiz ifadeler grubu iki alt türe ayrılmıştır. Belirsiz referansların ilk grubu, tüm derlemde en sık rastlanan belirsiz ifade olan *şey* (*AF*=2093, *RF*=12403.11) gibi spesifik olmayan varlıkları ifade eden belirsiz referanslardır. Bu ilk grupta, şey'e odaklanılarak belirsiz ifadelerin göndergeleri araştırılmıştır. Analiz, şey'in göndergesinin aynı ifadede, yakın bağlamında, genişletilmiş bağlamda bulunabileceğini veya göndergenin hiç bulunmayabileceğini göstermiştir. Tüm bu durumlarda, konuşmacıların paylaştığı 'ortak bilgi' nedeniyle etkileşim bozulmamıştır. İkinci grup, *insan* (*AF*=21, *RF*=124.45) ve *adam* (*AF*=5, *RF*=29.63) belirteçlerinin tanımlandığı genel gönderimlerdir. Bu belirteç grubunda yer alan *insan* sözcüğü, ilişkisel yönetim alanındaki işlevleri açısından tartışılmıştır. Sonuçlar, insan'ın jenerik referansının konuşmacılar tarafından genellikle olmayan bir diğerinin davranışlarına yönelik kişisel ve genellikle değerlendirici bir görüş iletmek için kullanıldığını göstermiştir.

Belirsiz eklentiler ise, belirteçler yaklaştırıcılar ve genel genişleticiler olarak gruplandırılmıştır. Yaklaştırıcılar, *neredeyse* (*AF*=22, *RF*=130.37) gibi nicelik veya durumların yaklaşık bir tahminini belirtmek için kullanılırken, bu belirsizlik kategorisine hakim olan grup genel genişleticiler (İng. general extenders) olmuştur (belirteçlerin %98'i genel genişletici olarak kodlanmıştır). Bulgular, Türkçe için genel genişleticilerin hem *vesaire* (*AF*=3, *RF*=17.77) gibi sözcüksel hem de *m- ikileme işaretleyicisi* (*AF*=16, *RF*=94.81) gibi ekler halinde kullanılabildiğini ortaya çıkarmıştır. Bu kullanımlarda Türkçe gençlik diline özgü çeşitli kullanım şekilleri gözlemlenmiştir. Bunlardan biri, konuşmacıların Türkçe'ye özgü bir morfolojik kural olan *m- ile ikileme oluşturma* tekniğini İngilizce kelimelere uygulayarak genel genişleticiler oluşturduklarını ortaya koymuştur. Genel genişleticiler yaratmanın bu yenilikçi kullanımı, konuşmacıların mevcut dilsel kaynaklarını tam olarak kullanan gençlik dilinin doğası gereği dinamik yapısını yansıtmaktadır.

Bulgular, yakın bağlamın belirsiz ifadelerin kullanımı üzerindeki etkisine dikkat çekmiştir. Bu amaçla, Biber ve diğerleri (2021) ile Egbert ve diğerlerinin (2021) resmi olmayan sözlü etkileşimdeki söylem birimlerinin (İng. discourse units) iletişimsel amaçlarına ilişkin taksonomisinden yararlanılmıştır. İncelenen belirsiz ifade, CoTY'de en sık rastlanan genel genişletici olan *f(a)lan* (*AF*=1468, *RF*=8699.36) olarak belirlenmiştir. Analiz, *f(a)lan*'ın taksonomideki tüm konuşma iletişimsel amaçlarında mevcut olduğunu göstermiştir. Bu iletişimsel amaçlar: (1) duruma bağlı yorum, (2) şakalaşma, (3) çatışmaya girme, (4) bir şeyleri anlamlandırma, (5) duygu ve değerlendirme paylaşımı, (6) tavsiye ve talimat verme, (7) geçmişi tanımlama veya açıklama, (8) geleceği tanımlama veya açıklama ve (9) tanımlama veya açıklama (zamandan bağımsız) olarak sıralanmaktadır. Belirsiz dilin samimi ve gayri resmi söylemlerde öne çıktığı yönündeki mevcut alanyazını (Channell, 1994; Clancy ve McCarthy, 2015; Clancy, 2016; Cutting, 2001) doğrular şekilde, genel genişletici *f(a)lan*, duygu ve değerlendirmelerin paylaşıldığı iletişimsel amaçlı söylem birimlerinde daha sık tespit edilmiştir (tüm amaçların %31'ine karşılık gelmektedir). Derlemdeki belirli iletişimsel amaç türlerinde *f(a)lan*'ın bir dizi farklı edimsel işlevi gözlemlenmiştir. En göze çarpanlar arasında, duygu ve değerlendirmelerin paylaşıldığı bölümlerde mevcut olan dedikodu konuşmalarında hafifletme işlevini, geçmişi tanımlama veya açıklama söylem birimlerinde *yeniden canlandırmayı* (İng. reenactment) birlikte inşa etmeyi ve geleceğe yönelik bir hipotezde dayanışmanın inşasını vurgulanmıştır.

Derlemdeki Etkileşim Belirleyicileri: Pekiştireçler

Pekiştireçler (İng. intensifiers), konuşmacılar tarafından iletilen mesajı abartmak veya azaltmak için kullanılan sözcüksel öğelerdir. Bunların tipik özellikleri olan üretkenlik, ifade gücü ve geri dönüşüm (Aijmer, 2018, 2020; Nevalainen ve Rissanen, 2002; Stoffel, 1901; Tagliamonte 2008) gençlik konuşmalarının yenilikçi doğasına çok uygundur. Bu çalışmanın amaçları doğrultusunda, tabu yoğunlaştırıcıların yanı sıra sıfat ve zarf yoğunlaştırıcıları da araştırmaya dahil edilmiştir. Biber ve diğerlerini (1999) takiben, derlemdeki yoğunlaştırıcıları kategorize etmek için kuvetlendiriciler (İng. amplifiers) ve düşürücüler (İng. downtoners) ikili kategorizasyonu kullanılmıştır.

CoTY'de 29 tür 2856 pekiştireç saptanmıştır. Bir sözlüksel öğenin anlamının belirli bir yönünün gücünü arttırmak için kullanılan kuvvetlendiriciler, pekiştireç türleri açısından daha zengin bulunmuştur (*n*=24) ve tüm pekiştireçlerin %93'ünü oluşturmaktadır. Bu grup içinde en sık rastlanan kuvvetlendirici, *çok* (*AF*=1705, *RF*=10103.82) olmuştur. Bu kuvvelendiricinin ardından gelen *bayağı* (*AF*=188, *RF*=1114.09) ise CoTY'deki anahtar sözcüklerden biri olarak da altı çizilen bir sözcük olarak öne çıkmaktadır. *Bayağı*'nın yanı sıra *aşırı* (*AF*=109, *RF*=645.93) da bir başka kuvvetlendiricidir. Anlamsal olarak benzer olmalarına rağmen, bu kuvvetlendiricilerin derlemdeki anlamsal bürünlerinin (İng. prosody) farklılık gösterdiği saptanmıştır. Bulgulara göre, *aşırı* bir kişinin olumsuz özelliklerinin altını çizmek için kullanılırken, *bayağı* olumlu özelliklerini vurgulamak için kullanılmıştır.

CoTY'deki kuvvetlendiriciler incelendiğinde, STD'nin sunduğu çağdaş genel konuşma Türkçesinde yer almayan tabu ve küfür ifadelerinin de (örn. *ana* sözcüğü ile üretilen küfür ifadeleri) kuvvetlendirici olarak gençlik dilinde kullanıldığı saptanmıştır. Buna ek olarak, İngilizce'den ödünç alınan *full* (*AF*=33, *RF*=195.56) ve artık Türkçede sıklıkla kullanılan bir sözcük olan *süper* sözcüğü gibi (*AF*=3, *RF*=17.78) gibi alıntı kelimelerin de kuvvetlendirici işlevlerle kullanıldığı saptanmıştır.

Kuvvetlendiricilerin tam tersi işelve sahip olan düşürücüler (İng. downtoners) iletilen mesajın gücünü azaltmayı hedefler. CoTY'de en sık rastlanan düşürücü *biraz(cık)* (*AF*=196, *RF*=1161.50) olarak saptanmış ve bunu *bir tık* (*AF*=26, *RF*=154.08) takip etmiştir. Sözlüksel öğelerin özgün anlamlarını kısmen ya da tamamen kaybederek pekiştireçlere dönüşüm süreci (Partington, 1993; Tagliamonte ve Roberts, 2005), *bir tık*

pekiştirecine odaklanılarak araştırılmıştır. Bu analizi yaparken, *bir tık*'ın Türkçe gençlik dili, genel konuşma Türkçesi (STD ve TUD sözlü alt derlemi) ve genel yazılı Türkçe (TUD) derlemlerindeki oluşumlarını ve edimlsel kullanımları artzamanlı bir yaklaşımla incelenmiştir. Bir örnekçe olarak *tık* STD'de mevcut olmasına rağmen (*AF*=12, *RF*=57.50), *bir tık* STD verilerinde mevcut değildir. STD'de *tık*, 'hızlı' ya da 'yok' anlamlarını iletmek için kullanılmıştır. Tamamlayıcı bir veri kaynağı olarak, TUD sözlü alt derlemi ise, *tık*'ın (*AF*=35, *RF*=34.52) STD'deki ile aynı anlamları sergilediğini, ancak sözlü alt derleminde hiçbir sonuç vermediğini göstermiştir. Yazılı TUD'daki sorgu ise, tıpkı CoTY'de olduğu gibi pekiştireç olarak kullanılan *bir tık* kullanımlarını ortaya çıkarmıştır. Bu kullanımlar (*n*=4) 2012 ve 2013 yıllarında yayınlanan ve Türkçenin resmi olmayan konuşma dilini yansıtan blog yazılarından elde edilen verilere aittir. Bulgular, *tık*'ın son on yılda sözlükselleşmeye uğrayarak pekiştireç *bir tık*'a dönüşmüş olabileceği savını destekler niteliktedir. Türkçede genç konuşucular tarafından konuşulan çağdaş bir argo pekiştireç olarak *bir tık*'ın yeni ortaya çıkan bu pekiştireç kullanımının geçmişi sanal alandaki genç yetişkin dilsel pratiklerinin derlem araçları ile izini sürebildiğimiz 2012 yılına kadar gitmektedir. Verilerin kapsamı sınırlı olsa dahi bu gözlem, dil değişimi sürecini araştırmak için derlem yöntemlerini kullanmanın olanaklarını orataya koymuştur.

SONUÇ VE ÖNERİLER

Mevcut literatür gençlik dilinin pragmatik/söylem işaretleyicileri, yoğunlaştırıcılar, rapor edilen konuşma, değişmez etiketler, küfür ve tabu sözcükleri gibi çeşitli sözcüksel özelliklerini vurgulamış olsa da, bu çalışma araştırma odağını anahtarlık analizi (İng. keyness analysis) üzerine temellendirmiştir. Başka bir deyişle, bu çalışma CoTY'de temsil edilen Türkçe gençlik konuşması için incelenecek ayırt edici özelliklerin kapsamının sınırlarını belirlemek üzere derlem odaklı bir yaklaşım benimsemiştir.

Türkçe gençlik dili üzerine ne mevcut bir derlem bulunmaktadır ne de alanyazında daha önce yapılmış bir derlem çalışmasına rastlanmıştır. Bu çalışma gençlik dili ve derlem dilbilimi kesişiminde gelecekteki çalışmalar için sağlam bir zemin oluşturmayı amaçlamıştır. Gençlik dili keşfedilecek zengin bir veri sunsa da, 18 yaş altı katılımcılara ulaşmanın zorlukları, konuşmacıların özel alanında doğal olarak gerçekleşen ve spontane konuşma verilerinin elde edilmesi ve verilerin sistematik dokümantasyonu ve analizi için kullanılan metodolojilerin azlığı, bugüne kadar Türk dilbilimi alanında

gençlik dilinin görünmez kalmasına neden olmuştur. Bu çalışma ile, metodolojinin şeffaflığını derlem yöntemleri ile sağlayarak ve veri toplama araç ve prosedürlerini diğer araştırmacılar için erişilebilir hale getirerek gelecek çalışmalar için bir yol haritası sunmaktadır. Bu çalışmanın da savunduğu ve uyguladığı şekilde, açık bilim uygulamalarının önceliklendirilmesi ve artırılması yoluyla, dilbilimde ortak çalışma modellerinin gençlerin dil verilerini daha görünür hale getireceği umulmaktadır.

Mevcut gençlik dili alanyazınında İngilizce dili üzerine yapılan araştırmaların baskınlığı devam etmektedir. Bu nedenle, bugüne kadar birçok araştırmacı, gençlerin dilsel pratiklerine ilişkin yürütülen çalışmalarda diller arası karşılaştırmaların eksikliği ve bu tür araştırmalara duyulan ihtiyacın altını çizmiştir. Bu anlamda, gençlik dili çalışmalarında derlem yaklaşımının benimsenmesi bu çağrılara cevap niteliğindedir. Derlem dilbilimsel çalışmalar bir hedef dilin eşzamanlı ve artzamanlı analizlerine de olanak tanımakta ve bir dilin farklı kayıtlarının veya zaman dilimlerinin derlemlerinin kullanılması, bir dildeki dilsel çeşitliliğin ve yenilik örneklerinin izini sürmek için sağlam kanıtlar sağlamaktadır. Buna ek olarak, gençlik derlemlerinin geliştirilmesinin birinci dil eğitiminin yanı sıra yabancı dil ve ikinci dil pedagojisine katkısı büyüktür. Gençlik dili derlemi bulguları; dil öğrenme ve öğretme süreçlerine dahil edilerek, her türden dil derlemi beceri geliştirme, müfredat ve materyal tasarımı için kapsamlı fırsatlar sunacaktır.

N. THESIS PERMISSION FORM / TEZ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Efeoğlu Özcan
Adı / Name : Esranur
Bölümü / Department : İngiliz Dili Öğretimi / English Language Teaching

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English): **THE CORPUS OF TURKISH YOUTH LANGUAGE (COTY): THE COMPILATION AND INTERACTIONAL DYNAMICS OF A SPOKEN CORPUS** / TÜRKÇE GENÇLİK DİLİ DERLEMİ (COTY): DERLEM OLUŞTURMA VE SÖZLÜ BİR DERLEMİN ETKİLEŞİMSEL DİNAMİKLERİ

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☐ **Doktora** / PhD ☒

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years. *** ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. ***** ☐

***** *Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature ..... **Tarih** / Date .....
(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)
(Library submission date. Please fill out by hand.)

Tezin son sayfasıdır. / This is the last page of the thesis/dissertation.