

ANGLICA

An International Journal of English Studies

32/2 2023

EDITORS

Marzena Sokołowska-Paryż [m.a.sokolowska-paryz@uw.edu.pl]
Anna Wojtyś [a.wojtys@uw.edu.pl]

ASSOCIATE EDITORS

Silvia Bruti [silvia.bruti@unipi.it]
Lourdes López Ropero [lourdes.lopez@ua.es]
Martin Löschnigg [martin.loeschnigg@uni-graz.at]
Jerzy Nykiel [jerzy.nykiel@uib.no]

ASSISTANT EDITORS

Magdalena Kizeweter [m.kizeweter@uw.edu.pl]
Dominika Lewandowska-Rodak [dominika.lewandowska@o2.pl]
Bartosz Lutostański [b.lutostanski@uw.edu.pl]
Przemysław Uściński [przemek.u@hotmail.com]

ENGLISH LANGUAGE EDITOR

Barry Keane [bkeane@uw.edu.pl]

ADVISORY BOARD

Michael Bilynsky, University of Lviv
Andrzej Bogusławski, University of Warsaw
Mirosława Buchholtz, Nicolaus Copernicus University, Toruń
Jan Čermák, Charles University, Prague
Edwin Duncan, Towson University
Jacek Fabiszak, Adam Mickiewicz University, Poznań
Elżbieta Foeller-Pituch, Northwestern University, Evanston-Chicago
Piotr Gąsiorowski, Adam Mickiewicz University, Poznań
Keith Hanley, Lancaster University
Andrea Herrera, University of Colorado, Colorado Springs
Christopher Knight, University of Montana
Marcin Krygier, Adam Mickiewicz University, Poznań
Krystyna Kujawińska-Courtney, University of Łódź
Brian Lowrey, Université de Picardie Jules Verne, Amiens
Zbigniew Mazur, Maria Curie-Skłodowska University, Lublin
Rafał Molencik, University of Silesia, Sosnowiec
John G. Newman, University of Texas Rio Grande Valley
Jerzy Rubach, University of Iowa
Piotr Ruskiewicz, Pedagogical University, Cracow
Krystyna Stamirowska, Jagiellonian University, Cracow
Merja Stenroos, University of Stavanger
Jeremy Tambling, University of Manchester
Peter de Voogd, University of Utrecht
Anna Walczuk, Jagiellonian University, Cracow
Jean Ward, University of Gdańsk
Jerzy Wełna, University of Warsaw
Florian Zappe, independent scholar

GUEST REVIEWERS

Radosław Dylewski, Adam Mickiewicz University in Poznań
Joanna Esquibel, Æ Academic Publishing, independent scholar
Marta Falkowska, University of Warsaw
Johan Franzon, University of Helsinki
Janet Fuller, University of Groningen
Daniel Karczewski, University of Białystok
Monika Konert-Panek, University of Warsaw
Aniela Korzeniowska, University of Warsaw
Karolina Krawczak-Glynn, Adam Mickiewicz University in Poznań
Katarzyna Matuszak, Poznań University of Technology
Derrick McClure, University of Aberdeen
Agnieszka Piskorska, University of Warsaw
Anna Rędzich-Korkuz, University of Warsaw
Maciej Rosiński, University of Warsaw
Javier Ruano García, University of Salamanca
Paweł Rydzewski, University of Warsaw
Gjertrud Stenbrenden, Inland Norway University of Applied Sciences
Izabela Szymańska, University of Warsaw
Ingrida Tatolytė, Vilnius University
Marcia Veirano Pinto, Federal University of São Paulo
Magdalena Walenta, University of Warsaw
Raffaele Zago, University of Catania
Joanna Zaleska, Humboldt University of Berlin



UNIVERSITY
OF WARSAW

Anglica An International Journal of English Studies

ISSN 0860-5734

www.anglica-journal.com

DOI: 10.7311/Anglica/32.2

Publisher:

Institute of English Studies University of Warsaw

ul. Dobra 55

00-312 Warszawa

Nakład: 30 egz.

Copyright 2023 by Institute of English Studies University of Warsaw

All right reserved

Typesetting:

Tomasz Gut

Cover design:

Tomasz Gut

Printing and binding:

Sowa – Druk na życzenie

www.sowadruk.pl

+48 22 431 81 40

TABLE OF CONTENTS

Artur Kijak

On the Coronal Palatalization in Early Modern and Present-Day English . . . 5

Marie Flesh

“Dude” and “Dudette”, “Bro” and “Sis”: A Diachronic Study of Four
Address Terms in the *TV Corpus* 23

Marta Carretero

The Expression of Epistemicity in British Internet Discussion Forums
in Contrast with Newspaper Opinion Articles and Political Speeches 43

Mersina Mujagić

Figurative Conceptualizations of Nations, Countries, and Institutions
in Newspaper Articles on Migration 71

Nazi Iritspukhova

Metaphor and English Promotional Tourism Discourse: Systematic-narrative
Hybrid Literature Review and Future Research Areas 93

Hasnaa Hasan Sultan Abdelreheem

The Key English Pronunciation Difficulties for Egyptian EFL Learners . . . 115

Miłosz Marcjanik

EMI in Higher Education: Current Challenges 135

Katarzyna Jaworska-Biskup

Food Symbolism and Imagery in the Polish Translations of William
Shakespeare’s *The Merry Wives of Windsor* 135

Marie Flesch

 <https://orcid.org/0000-0003-2139-1517>

University of Lorraine

“Dude” and “Dudette”, “Bro” and “Sis”: A Diachronic Study of Four Address Terms in the *TV Corpus*

Abstract: This corpus linguistics study offers a diachronic perspective on masculine and feminine address terms by analyzing the frequencies of the pairs “dude”/“dudette” and “bro”/“sis” in the 325 million-word *TV Corpus*. Results show an increase of the frequency of “dude” and “bro”, a quasi-absence of “dudette”, and a stable but low frequency of “sis”. They suggest that “dude” and “bro” have taken on generic meanings, while the usage of “sis” remains kinship-specific. They also show that familiarizers are more frequent in American English than in British and Canadian English, and that their frequency in the *TV Corpus* is genre-dependent, with animated series and reality television shows being more conducive to their use.

Keywords: corpus linguistics, diachrony, address terms, gender

1. Introduction

TV dialogue is of particular interest to linguists as it may reflect or even influence ongoing language change (Heyd 2010). Most studies have focused on relatively small corpora compiled by the researchers themselves, such as the *Sydney Corpus of Television Dialogue* (Bednarek 2018), or have focused on a specific television show, for example *Friends* in Heyd (2010) and Tagliamonte (2005). The *TV Corpus*, created at Brigham Young University and released in February 2019 (Davies 2021), has opened new perspectives thanks to its sheer size (more than 325 million words), its diachronic span (from 1950 to 2017), and its rich metadata (country, date, genre, runtime). It can be combined with the IMDb database, allowing for further analyses. It can be used to conduct diachronic analyses which explore the changes in the frequency of words or lexical bundles year after year, as well as multivariate

analyses. It is, according to its homepage, a “great resource to look at very informal language”, which offers “unparalleled insight into variation in English” (*The TV Corpus* n.d.). The corpus has been used by recent studies, which have explored variation across genres (Jucker 2021), diachronic variation (Werner 2021; Landert, Säily, and Hämäläinen 2023), and geographic variation (Kang 2022; Hirota and Brinton 2023). This article takes advantage of the size and of the annotation of the *TV Corpus* to offer a diachronic perspective on the use of four address terms. As previous studies of speech, online texts, and TV and movie dialogue have focused on masculine forms which, for some of them (“guys”, “dude”), have taken on a generic meaning, we decided to also include their feminine counterparts: “dudette” for “dude”, and “sis” for “bro”. We provide a diachronic account of the frequency of each term, as well as inferential analyses that compare the use of address terms in American, British, and Canadian English, and in four TV genres: scripted shows, animated shows, documentaries, and reality television. The study is based on concordances extracted from the free version of the *TV Corpus* and merged with the metadata spreadsheet provided by the English-corpora.org website. Carried out with the R software, the analyses rely mainly on the peaks and troughs method (Brezina 2018) and multiple regression models.

2. Address terms

“Guys”, “honey”, “buddy”, “dude”, “dudette”, “bro”, and “sis” are familiarizers and belong to one of the five semantic categories of vocatives described by Leech (1999). According to Leech, they have three functions: to attract attention, select an addressee, and enhance the familiarity between the speaker and their addressee. Hill (1994) retraces the history of “dude”, from its earliest recorded use to refer to a person in England, as a “dudesman” or scarecrow wearing old rags. He describes the various shifts in the meaning of the term, which was used to refer to a “dandy” or a “well-dressed man” in late 19th century North America, and then lost its pejorative connotation in the middle of the 20th century by entering the language of Mexican American and African American subcultures. It started to be used by white surfers in the 1960s, which helped spread the term across all socioeconomic lines with the meaning “guy”. Hill suggests that, starting from the 1950s, TV had a profound effect on the rise of the address term. He also points out that the meaning of the term widened in the 1980s. It became an exclamation of delight or affection and an expression of disappointment. In the late 19th century, feminine forms of “dude” emerged; “dudine”, “dudette”, and “dudenette” were used in British and American magazines. Even if for Hill (1994) it may have been an early example of “artificial slang”, they were still used in the 1930s to refer to female guests in dude ranches, which hosted tourists (Johnson 2012). Recent studies suggest that “dude” may be becoming a gender-neutral address term. Kiesling (2004), for example, basing his

research on surveys and ethnographic data, suggests that “dude” indexes a stance of “cool” solidarity and is being increasingly used by women. He also found that, when used by women, “dude” may express a stance of distance or non-intimacy from a man. A self-report survey conducted in the United Kingdom found that the familiarizer is used as frequently by men as by women and nonbinary individuals, and that it is thought of as gender-neutral by younger respondents (Pastorino 2022).

“Bro”, the abbreviated form of “brother”, is attested from the 17th century (“Bro” n.d.). Its use as an address term may come from African American Vernacular English, where it refers to either a “man” or a “Black man” (Schwiegershausen 2013). “Bro” has risen in popularity across race lines and is now associated with “bro culture”, a specific type of masculinity exemplified by white college fraternity, and present in sports, country clubs, the military, and male-dominated professions (Jones 2017). The word has been used to create neologisms such as “bromance”, “bro-hug”, “bro-down” (“Bro Culture”). “Bro” has not attracted the attention of linguists as much as “dude”; Urichuk and Loureiro-Rodríguez (2019), who conducted a self-report survey in Canada, calling masculine vocatives “brocatives”, found that “bro” and its variants “brah” and “bruh” are more gendered than “dude”, with few women reporting being likely to choose it to address another woman. Unlike masculine address terms, feminine familiarizers have not been the focus of many linguistic studies. It may be due to their scarcity in corpora: they rarely develop generic meanings, while masculine terms such as “guys” or “dude” do (Clancy 1999). However, due to the dearth of research about terms such as “woman”, “sis” and “girl” in large corpora, we do not know much about their frequency and the social meanings they index. The frequency of address terms varies across languages. They are, for example, particularly frequent in Spanish (Kleinknecht 2013). Using a corpus-based approach to compare their use in British and American English conversation, Leech (1999) found familiarizers to be more frequent in American English, noting that British English relies more on kinship terms.

3. Corpus studies of TV dialogue

Studies of TV and movie dialogue have investigated a wide range of topics, such as issues related to dubbing and subtitling (Baumgarten 2008; Baños, Bruti, and Zanotti 2013; Lu 2023), gender roles and stereotypes (Bednarek 2015; Gregori-Signes 2017; Csomay and Young 2021; Li, Liu, and Liu 2022), or the way language is used to create characters (Reichelt 2018; Bednarek 2023). Many have compared scripted dialogue to unscripted language to see to what extent it differs from natural conversations, using various methods such as Biber’s multidimensional approach (Quaglio 2008), frequency lists (Bednarek 2011), and n-grams (Bednarek 2012; Levshina 2017). They have found that TV dialogue is a close approximation of informal speech, but is more emotional and less vague than spontaneous speech

(Bednarek, 2012; Levshina, 2017). Heyd (2010) describes TV dialogue as “staged orality”: like spontaneous speech, TV dialogue is oral, but it is designed by a team of writers and producers. Formentelli (2014) also highlights the dual nature of TV dialogue, specifically in the use of vocatives. On the one hand, it reproduces the interpersonal function and sociolinguistic variation of spontaneous speech. On the other hand, it uses sophisticated address strategies that allow for “authorial expressivity” (Formentelli 2014, 53). Tagliamonte (2005), who analyzed the use of intensifiers in the TV series *Friends*, points out that the series’ dialogues reflect spontaneous language use, as it exhibits the same overall rate and distribution of intensification as similar corpora of the English language. She also suggests that TV dialogue is more innovative than the spontaneous speech of the general population in the use of the intensifier “so”, “the new favorite in American English”, and that it provides a “kind of preview of mainstream language” (Tagliamonte 2005, 296).

Corpora of TV and movie dialogue also lend themselves to diachronic studies. Some have used dialogue from a single series that spanned many years, such as *Star Trek* (Csomay and Young 2021), small corpora of movie dialogue (Forchini 2013), or parallel corpora (Formentelli 2014). The *TV Corpus* and the *Movie Corpus*, released in 2019, spawned a new wave of studies looking at the diachronic development of words, such as the use of the verbs “speak” and “talk” (Kang 2022), the suffix “-ish” (Eitelmann and Haumann 2023), and polite expressions and swear words (Jucker and Landert 2023). Landert, Säily, and Hämäläinen (2023) used the *TV Corpus* to identify words that appeared on TV earlier than their first date of attestation in dictionaries, noticing some encoding issues in the corpus and errors in the metadata, and emphasizing the need for manual verification of concordances.

Many studies of TV and movie dialogue have looked at features of orality, such as the expression “me likey” (Rodríguez-Abruñeiras 2022), the pragmatic marker “you bet” (Hirota and Brinton 2023), contractions, interjections, and discourse markers (Jucker 2021). Address terms have been the focus of several studies. Bruti and Perego (2010) analyzed the distribution of vocatives in seven movies and four TV episodes, coming to the conclusion that their frequency is genre-dependent. Heyd (2010) studied a corpus of the TV show *Friends*, focusing on the emergence of “you guys” as a second-person plural pronoun. Comparing the *Pavia Corpus of Film Dialogue* and the *Longman Spoken and Written Corpus*, Formentelli (2014) found that names, kinship terms, familiarizers, endearments, and insults were four times more frequent in film dialogue than in spontaneous speech. Quaglio (2009), who compared the language of the *Friends* TV show to the American conversation subcorpus of the *Longman Grammar Corpus*, found “guys”, “man”, “dude” and “buddy” to be a lot more frequent in the TV show than in natural conversation. Other studies of address terms have adopted a diachronic approach. In a small corpus of movies, Forchini (2013) noted a rise over time in the frequency of the familiarizers “man”, “guys”, “buddy” and “dude”. Werner (2021) looked at the frequency of “dude”, “buddy” and “bro” (among other terms) in the American portion of the

TV Corpus, noting their increasing presence in the TV dialogue and highlighting the rise in familiarity of American English.

4. Methods

4.1 The *TV Corpus*

The *TV Corpus* was created at Brigham Young University using texts taken from the OpenSubtitles collection, which were matched with IMDb pages providing metadata for each TV show (*The TV Corpus* n.d.). According to the metadata provided on the corpus page (https://www.english-corpora.org/tv/files/sources_tv.zip), it contains 75,804 episodes from 2,988 series that aired between 1950 and 2017. Many series feature more than once in the corpus, as they span several years. The corpus is a sample of TV series: not all episodes of the series are included. The corpus contains on average 110.18 series per year ($SD = 165.30$), with a minimum of 1 series in 1950, and a maximum of 625 series in 2016. It consists of 326,201,276 tokens (or words), with a mean number of tokens of 4,797,078 per year ($SD = 6,958,599$), a minimum of 9,484 tokens in 1950, and a maximum of 25,077,851 in 2016. Figure 1 presents the number of tokens per year in the corpus. It shows that there is relatively little data in the first four decades of the corpus and that there is a tenfold increase in yearly corpus size between the early 1990s and the mid-2010s.

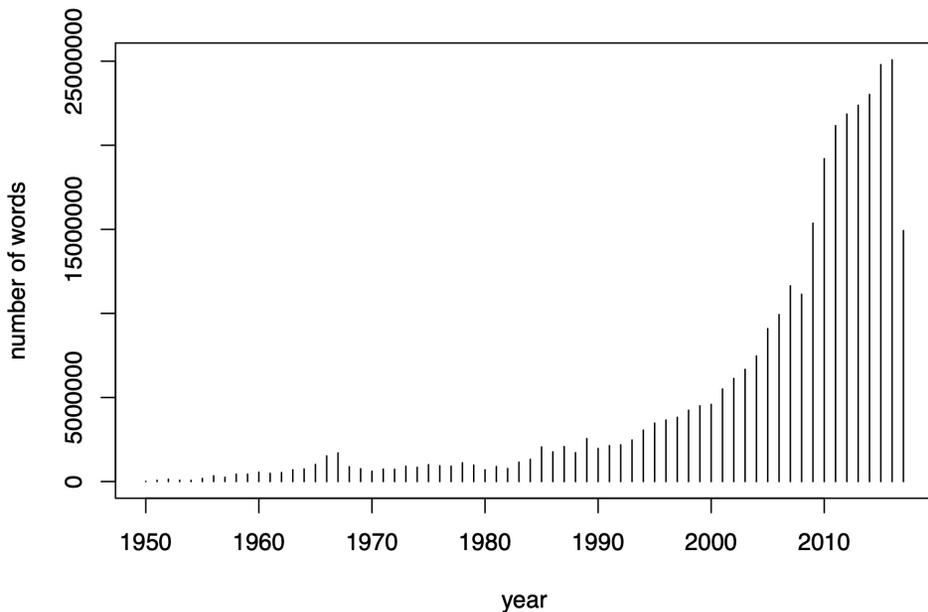


Fig. 1. Number of tokens per year in the *TV Corpus*

4.2 Data collection and processing

We used the free version of the corpus, which is available on the website English-corpora.org (*The TV Corpus* n.d.). Concordances were generated in January 2022 using the online interface, with four search terms: “dude”, “dudette”, “bro”, and “sis”. Concordance lines were then pasted into a .csv file, with the metadata provided by the online version of the corpus (year and title). We downloaded the .xlsx file provided on the English-corpora.org website which offers additional metadata associated with the texts in the corpus: number of words per series episode, language, country, genre, IMDb number, and runtime. The concordances (available on OSF: https://osf.io/rb2ez/?view_only=483c7270e7f048e386c7f5fa9fb6168b) were then merged with the metadata and analyzed with the software program R. All the frequency analyses presented in the article were carried out with R. We did not use the “chart” tool provided by the *TV Corpus* interface. Four series featured in the concordance lines and not in the metadata and were thus eliminated from the analysis. This explains the small discrepancies between the overall frequency of the four address terms provided by the online corpora and the results presented in this article. We manually examined the concordance lines for each search term and classified them into four categories: reference terms, address terms, verbs, and metalanguage:

- The “reference term” category contains occurrences where “dude”, “dudette”, “bro”, and “sis” are used to refer to a person or a thing, as in “You bastards! You telling on our sis!” (*Skins* 2007) and “Just, like, a handsome dude in a classic sense” (*Happy Endings* 2011).
- The “address term” category comprises instances when the terms are used to address someone, as in: “I told you dude it’s your citrus intake” (*Gilmore Girls* 2003) or “What does that say, bro?” (*The Voice* 2013). As address terms can sometimes develop into interjections and discourse markers (Sonnenhauser and Noel Aziz Hanna 2013), this category also very likely includes instances where the familiarizers are used as interjections. Because it is difficult to differentiate between both usages when looking only at the text, we did not create a special category for interjections.
- In the “metalanguage” category, the usage of the terms is discussed by the characters, for example as not being appropriate to address a person, as in: “Dude, bros don’t even use ‘bro’” (*Scrubs* 2006) and “I only call people in my crew ‘Bro’. Bro and Bra, two totally different things” (*Kath and Kim* 2009).
- The “verb” category includes instances where the nouns become verbs through conversion, as in: “Segal and I bro’d out like crazy” (*Greek* 2011) and “Don’t ‘dude’ me” (*The New Adventures of Old Christine* 2008).

In some cases, it was not possible, based on the text of the dialogue alone, to distinguish between address and reference term, as in “I know that dude!” (*Pinky and the Brain* 1998) or “How you gonna handle this dude?” (*Dark Angel* 2001).

These occurrences were categorized as “unknown”. We eliminated instances where the search terms are part of names of movies, bands, brands, and organizations, like in the following examples:

“Oh, that Ashton Kutcher movie. Right. He was so funny in ‘dude, where’s my car?’” (*Switched at Birth* 2013).

“Hey, George, what do you like better, the Bro or the Mansiere?” (*Seinfeld* 1995)

We then tallied the raw frequency of each term not per episode or series, but per “series-year”. We created this text unit by concatenating the name of each series with each year it is included in the corpus. For example, the series-year *The Sopranos*-1999 comprises all the episodes of *The Sopranos* that came out in 1999 and are included in the corpus. *The Sopranos*-2000 contains all the episodes of *The Sopranos* that came out in 2000 and are included in the corpus, and so on. We had to proceed this way because the online interface of the *TV Corpus* only provides the name and the year of a concordance line, and not a unique identifier such as the episode name. We then computed the relative frequency of the terms for each of the 7,492 “series-years” we identified.

5. Results

5.1 “Dude”

42,593 occurrences of “dude” were found in the corpus, including 34,204 address terms, 8,247 reference terms, 4 verbs, 17 instances of metalanguage, and 121 unknown. The address term appears 10.49 times per 100,000 tokens, and the reference term 2.53 times. Figure 2 shows the relative frequency of “dude” as an address and reference term in the corpus, per year. “Dude” appears for the first time as a reference term in 1960. It was not used much until the mid 1980s. It was not used as an address term at all in 23 years of the corpus, mostly in the 1950s and 1960s, but also in the 1970s and 1980s. It was found in 3,466 (46.26%) series-year. The graph reveals an increase in frequency, starting in the mid-1980s, as well as two peaks: the first in the early 1990s, and the second a little before 2010. The frequency of “dude” more than doubles between 1995 and 2010, and reaches its maximum in 2010, with 17.70 occurrences per 100,000 tokens. The increase in frequency seems to mainly affect the use of “dude” as an address term, with a more modest increase in the frequency of the reference term.

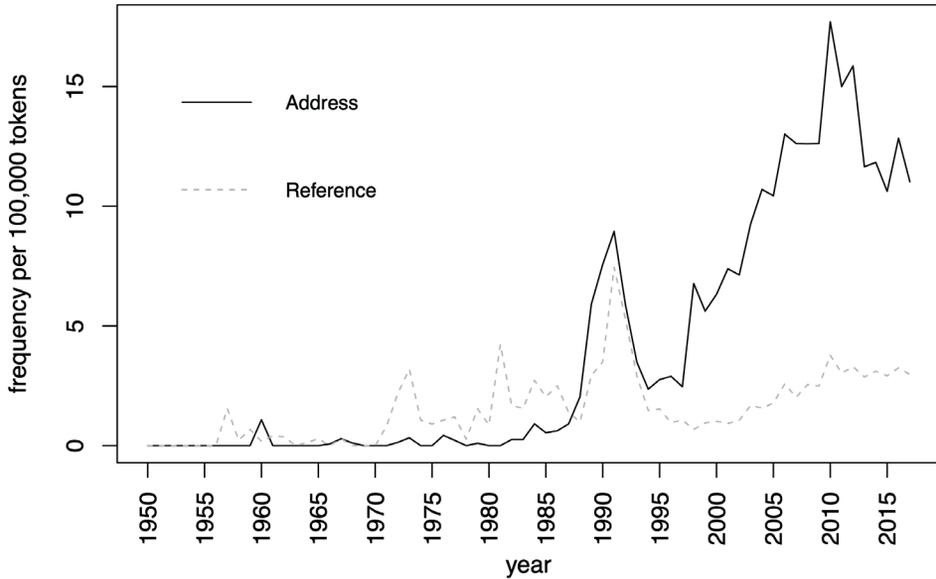


Fig. 2. Frequency of “dude” in the *TV Corpus*

When looking at this graph and at the other graphs presented in this article, it is important to remember that corpus size varies greatly, with later years of the corpus containing a lot more data than the first eight decades. Thus, individual series-year have a larger impact on smaller subcorpora. The first peak of the graph may in part be explained by the animated series *Teenage Mutant Ninja Turtles*. In 1991, for instance, 79.06% (151) of all occurrences of “dude” (191) in the year appear in this series, which represents 20.62% (103,528 tokens) of the data in the corpus for that year (2,135,182 tokens). The series, which features in the corpus from 1987 to 2016, consistently exhibits a large frequency of “dude” starting from 1989. However, its impact lessens in later years of the corpus, where it accounts for a smaller proportion of the data. In 2014, for example, “dude” has a relative frequency of 14.73 per 100,000 tokens in the series but the series only accounts for 0.20% (46,054 tokens) of the dialogue included that year (23,022,413 tokens). To distinguish between statistically significant and non-statistically significant variation, we used the “peaks and troughs” method. Described by Brezina (2018), it applies a Generalized Additive Model (GAM) to frequency data over time and is used to highlight significant outliers in a diachronic corpus. The resulting graph is presented in Figure 3. It shows that the frequency of “dude” is stable until the early 1980s and that it starts increasing, with a small dip in the mid-1990s and a dramatic increase that culminates in 2010.

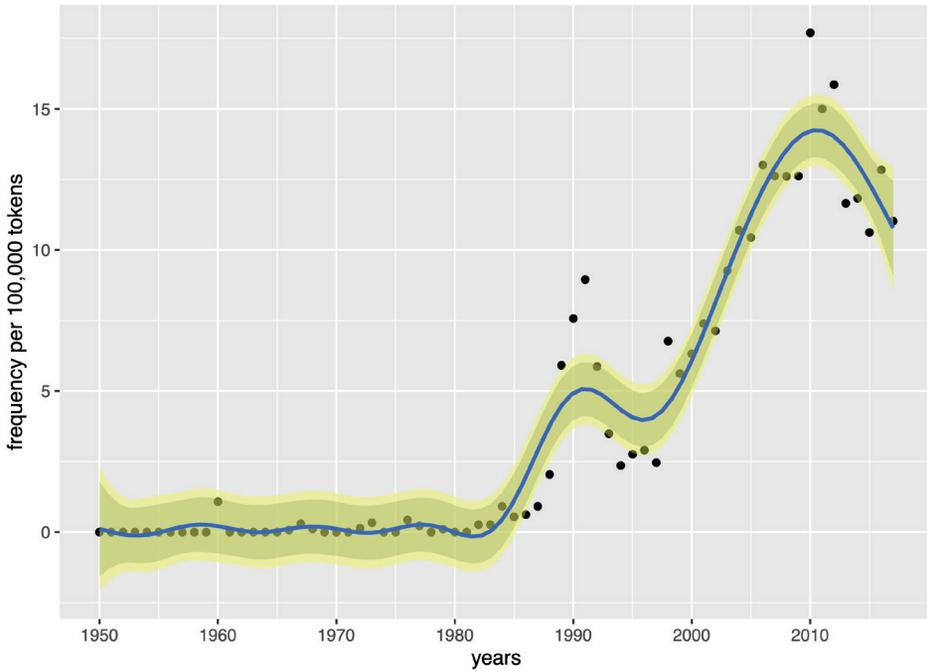


Fig. 3. Peaks and troughs analysis of the address term “dude” in the *TV Corpus*

5.2 “Bro”

“Bro” appears 11,173 times in the corpus, including 10,451 occurrences as an address term, 642 as a reference term, 44 in metalanguage, and 35 as a verb. It occurs 3.2 times per 100,000 tokens as an address term and 0.2 times as a reference term. It was used in 2,283 series-year (30.47% of series-year in the corpus). It appears for the first time as an address term in 1973 and as a reference term in 1985. It does not feature as an address term in 29 years of the corpus. While the frequency of the address term increases with time, the frequency of the reference term is close to zero in all years of the corpus (Figure 4). Peaks and troughs analysis (Figure 5) reveals that the frequency of the term does increase with time, starting in the mid-1970s. It also shows a number of outliers, corresponding to the points outside of the yellow area.

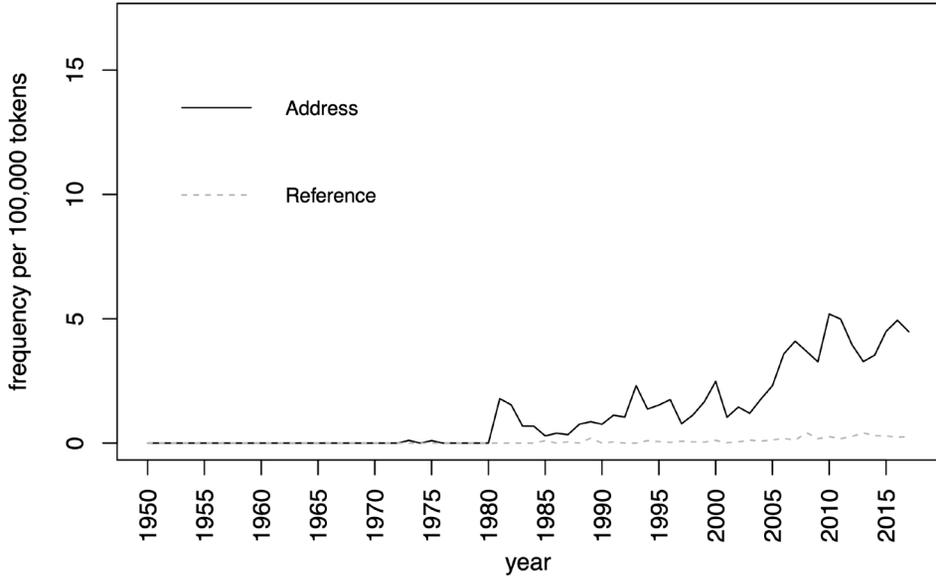


Fig. 4. Frequency of “bro” in the *TV Corpus*

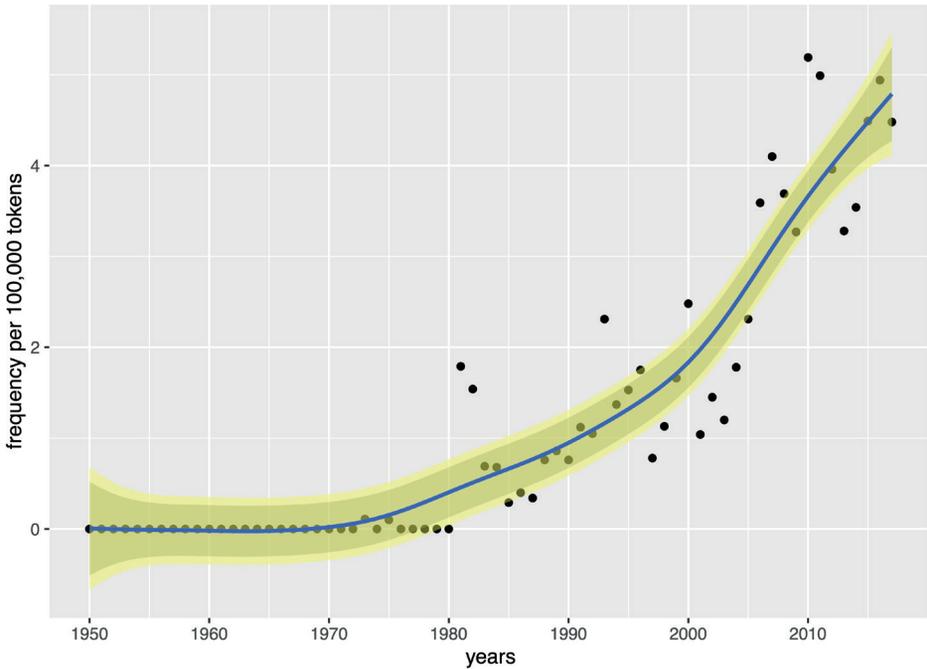


Fig. 5. Peaks and troughs analysis of the frequency of the address term “bro” in the *TV Corpus*

5.3 “Dudette”

We counted only 81 occurrences of “dudette” in the corpus, including 58 address terms and 20 reference terms. Consequently, the relative frequency of the term is close to zero, with 0.02 occurrence per 100,000 as an address term, and 0.01 as a reference term. Figure 6 shows its frequency through time in the corpus; the term was employed for the first time in 1981, as a reference term, in the series *Magnum, P.I.* “Dudette” only features in 14 series; one animated series, *Teenage Mutant Ninja Turtles*, is responsible for the bulk of occurrences, with 63 occurrences in total (14 as a reference term, and 49 as an address term), or 77.78% of all occurrences of the term in the corpus. These occurrences account for the peak represented in the graph, around the years 1991 and 1992. After this small spike, the frequency of the term drops and stays close to zero. Since “dudette” is quasi-absent from the corpus, we did not conduct further analysis of the term.

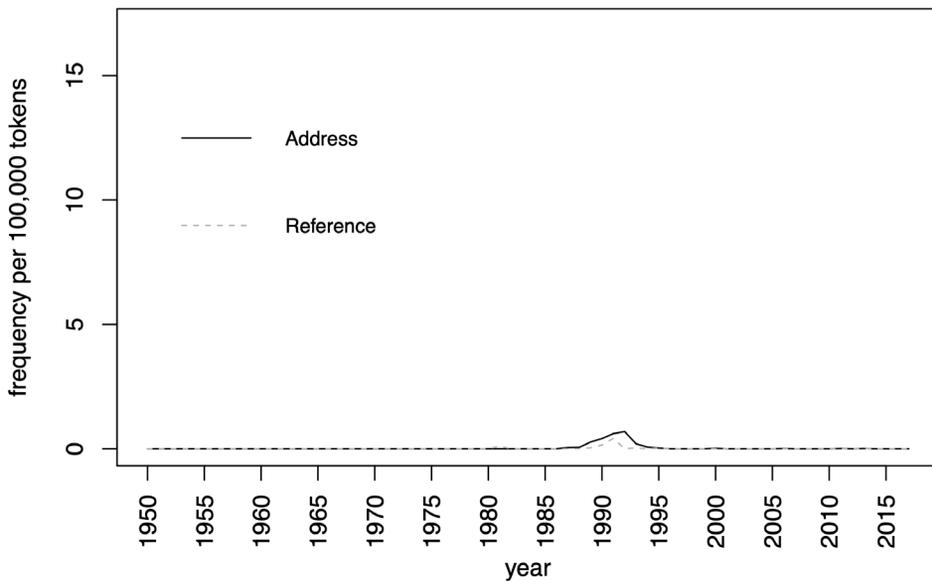


Fig. 6. Frequency of “dudette” in the *TV Corpus*

5.4 “Sis”

2,019 occurrences of “sis” were found in the corpus: it features 1,722 times as an address term, 290 times as a reference term, once as a verb, and twice in metalanguage. Its overall frequency per 100,000 is 0.62 (0.53 as an address term, and 0.09 as a reference term). Figure 7 shows its frequency through time as an address and

reference term. After the first recorded use of “sis” in the corpus, in 1956, the graph does not reveal any clear pattern. It does not indicate an increase in frequency, but several small spikes. As an address term, “sis” occurs in 197 series-year or 2.63% of the series-years in the corpus. As no clear pattern emerged from the descriptive statistics, we did not perform peaks and troughs analysis for the term.

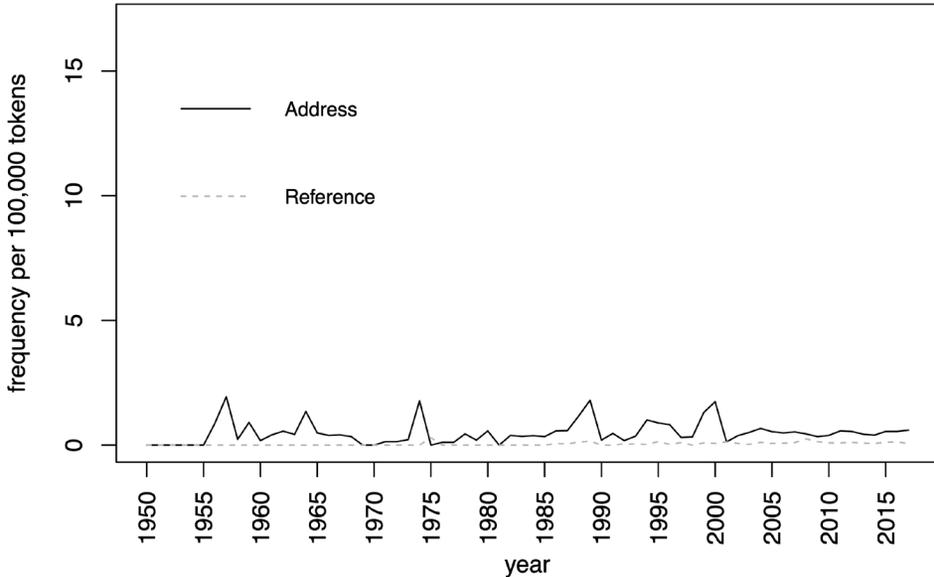


Fig. 7. Frequency of “sis” in the *TV Corpus*

5.5 All familiarizers

Figure 8 combines the line graphs corresponding to the frequency of each address term. It highlights the predominance of “dude” over the other three familiarizers. It also suggests that, even if “bro” is much less frequent than “dude”, its diachronic trajectory seems to parallel that of “dude”, with an increase starting in the early 1980s. “Sis”, on the other hand, does not follow this trend.

5.6 Analysis by country

Three subcorpora were created, corresponding to the series produced in the US, in the UK, and in Canada. The US corpus is by far the largest. It is almost five times bigger than the UK corpus, and 14 times larger than the Canadian corpus. Table 1 presents the relative frequency of the three address terms in the corpus per 100,000 tokens. It shows that all address terms are more frequent in the US subcorpus than in the other two subcorpora, with “dude” being about ten times more frequent than in the UK. The three familiarizers are the least frequent in the UK subcorpus.

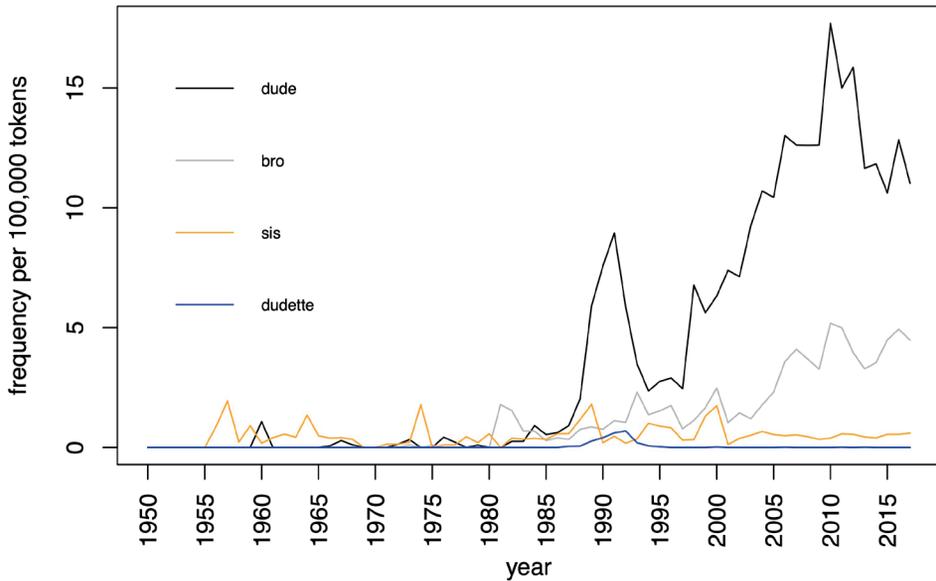


Fig. 8. Frequency of the four address terms in the *TV Corpus*

Table 1. Frequency of the familiarizers by country, per 100,000 tokens

	Tokens	Dude	Bro	Sis
US	243,674,852	12.77	3.84	0.55
UK	51,990,177	1.19	0.66	0.22
Canada	17,125,496	9.13	2.19	0.29

To find out if the differences between the three subcorpora are significant, we created negative binomial regression models, which are able to deal with the dispersion typical of corpus linguistics data (Hilbe 2011). The raw frequency of each address term is the dependent variable of each model; country and time were the predictors, and the number of tokens in each “series-year” was used as an offset. Since the Canadian subcorpus only contains very few series before the year 1995, we only considered data from 1995 to 2017. The results are presented in Table 2. Coefficients have been exponentialized (as is the case in the other models presented in this article). There is a positive and significant correlation between time and the frequency of “dude” and “bro”, meaning that it increases with time. For “sis”, however, the variable was not significant and was removed during the variable selection process, using the step() function in R. The models reveal significant differences between all countries, for all familiarizers. “Dude”, “bro”, and “sis” are significantly more frequent in the US subcorpus than in the other two subcorpora. The difference between the US subcorpus and the UK subcorpus is larger than that

between the US and the Canadian subcorpora. For example, the probability of the address term “dude” being used is 11.75 higher in the US corpus than in the UK corpus, and 9.15 times higher in the Canadian subcorpus than in the UK subcorpus. For “bro”, the effect size is smaller: according to the model, it is 5.68 times more frequent in the US than in the UK subcorpus, and 3.46 times more frequent in the Canadian subcorpus than in the UK subcorpus. “Sis” is twice as frequent in the US subcorpus than in the UK subcorpus; the probability of it occurring in the Canadian subcorpus is 1.81 higher than it is in the UK subcorpus. It is also significantly more frequent in the US corpus than in the Canadian subcorpus.

Table 2. Regression models with countries and time as the dependent variables, with coefficient and confidence intervals (* $p < 0.05$; ** $p < 0.01$)

	Dude	Bro	Sis
Intercept	0.000** (0.00, 0.00)	0.000** (0.00, 0.00)	0.000** (0.00, 0.00)
Canada	0.779* (0.64, 0.95)	0.609** (0.47, 0.79)	0.557** (0.36, 0.87)
UK	0.085** (0.07, 0.10)	0.176** (0.15, 0.21)	0.480** (0.36, 0.63)
Year	1.038** (1.03, 1.05)	1.076** (1.06, 1.09)	-

5.7 Analysis by genre

The *TV Corpus* metadata provides information about the genres of the series. Unfortunately, most series are classified as belonging to several genres, such as “Crime, Drama, Mystery” or “Action, Adventure, Drama”. For this analysis, we decided to use only four categories: “Animation”, “Documentary”, “Reality TV” and “Scripted series”, a category that includes all non-animated scripted shows in the database. This classification is based on the assumptions that the dialogue of reality television shows may be closer to authentic interactions than the dialogue of scripted series and that documentaries may contain less informal language than other genres. The category “Animation” was deemed relevant after our initial exploration of the corpus, which suggested that “dude” may be more frequently used in animated series, often geared towards children and teens. We used an automated method to extract data from the strings of genres provided in the corpus metadata: we extracted the first of the three genres under examination (“Animation”, “Documentary”, “Reality TV”), and classified all other series into the fourth umbrella category. To exclude the effect of country on our analysis, we only considered American texts. The relative frequencies of the familiarizers in the four subcorpora we created are presented in Table 3. It shows that “dude” and “bro” are more frequent in animated and in reality television series than in the other two categories. Animation also exhibits the highest frequency of “sis”. As we expected, the frequency of the four address terms is the lowest in documentaries.

Table 3. Frequency of the familiarizers by genre

	Tokens	Dude	Bro	Sis
Scripted	209,823,804	10.53	3.44	0.54
Animation	16,117,503	37.56	7.27	1.15
Reality TV	10,967,036	23.67	8.32	0.24
Documentary	6,766,509	5.62	0.7	0.06

We created a negative binomial regression model to determine if the differences in frequency are significant, with time and genre as independent variables. As no reality television series are included in the corpus before 2000, we only considered US data dating from 2000. The results show significant differences between all genres. “Dude” and “bro” are more frequent in animated series than in other genres. They are also more frequent in reality television series than in documentaries and scripted series (other than animated series). “Sis” is also more frequent in animated series than in any other genre and more frequent in scripted series than in reality television and in documentaries.

Table 4. Regression models with genre and time as the dependent variables, with coefficient and confidence intervals (*p<0.05; **p<0.01)

	Dude	Bro	Sis
Intercept	0.000** (0.00, 0.00)	0.000** (0.00, 0.00)	0.000** (0.00, 0.00)
Animation	2.464** (2.12, 2.88)	2.011** (1.66, 2.45)	2.188** (1.58, 3.06)
Documentary	0.391** (0.30, 0.51)	0.167** (0.11, 0.25)	0.111** (0.03, 0.29)
Reality TV	1.653** (1.33, 2.08)	1.550** (1.19, 2.05)	0.339** (0.18, 0.62)
Year	1.017** (1.00, 1.03)	1.056** (1.040, 1.072)	-

6. Conclusions

Our study shows that “dude” is the most frequent of the four terms we analyzed, and that its frequency has increased dramatically over the past three decades, especially as an address term. The fact that “dude” may be becoming more gender-neutral (Kiesling, 2004; Pastorino, 2022) may account for its rise in popularity: it is no longer a word used only by men, to talk to men. The fact that, in our data, several series that exhibit the highest relative frequency of “dude” are series that mostly feature female characters (*Broad City*, *Kath and Kim*, *Girl/Girl Scene*) seems to support this. Like many other masculine terms, such as “man” or “guys”, “dude” seems to have taken

on a generic meaning. “Dudette”, on the other hand, seems to have known the fate predicted by Hill (1994), who described it as “artificial slang”. Mostly used in one series, during a short period of time, it then disappeared from the corpus. Kiesling (2004) suggests that the lack of success of “dudette” has to do with the fact that it is a diminutive derivative of “dude” and has a negative connotation. The analysis of the pair “bro”/“sis” also reveals an asymmetry between the two address terms. Like “dude”, “bro” rises in popularity over the years. However, this increase in frequency only concerns the address term, and not the reference term. In the *TV Corpus*, “bro” is less frequent than “dude”, probably because the term indexes a specific type of masculinity. However, the term assumes a more generic meaning than “sis”. The clipping of “brother” can be used to address a family member or a man, while, we hypothesize, “sis” is primarily used to talk to a speaker’s sister. Again, looking at the series where “sis” is the most frequent supports our analysis. “Sis” is the most frequent in *Sonic Underground*, *Fantastic Four*, and *The Spoils of Babylon*, three series that depict fraternal relations (Sonic the Hedgehog and his sister Sonia, Lauoreighiya Samcake, Eric Jonrosh’s sister in *The Spoils of Babylon*, and Susan Storm and her brother Human Torch in *Fantastic Four*). Thus, we conclude that not only is “sis” gender-specific, but it may also be mainly kinship-specific. This may explain why the frequency of the address term has remained stable since its appearance in the corpus: it has not taken any additional meaning other than “sister” or, if it has, this other meaning remains rare in television texts.

The analysis of the four genres shows that animated series use familiarizers more frequently than scripted series, documentaries, and reality television series. Since animated series are often geared toward children and teenagers, this confirms the connection between “dude”, “bro” and teenage speech. Including reality television in our analysis was an attempt to find out if the frequency of address terms in scripted series reflects actual usage. It was based on the idea that since reality television shows are not scripted, they may reflect spontaneous speech more accurately. If we accept this hypothesis, then our results suggest that “dude” or “bro” may be more frequent in spontaneous interactions than in TV dialogue, since we found a significant difference. However, other studies such as Quaglio (2009) and Formentelli (2014) found that familiarizers are more frequent in TV dialogue than in spontaneous speech. Thus, it is possible that the high frequency of “dude” and “bro” in reality television texts may be due to the specific interactions of the reality television genre, which may not accurately reflect natural occurring conversations.

The comparison between the three varieties of English indicates, as we expected, that the familiarizers studied here are a feature of American English, and to a lesser extent of Canadian English. It confirms what Leech (1999) noted in a corpus of British and American conversations. Our analysis reveals a significant difference, but with a relatively small effect size, between Canadian and American English. This difference may be due to the nature of the texts included in each corpus. For instance, there may be more animated movies and reality

television shows in the American corpus, which could account for the difference. The increase in the frequency of “bro” and “dude” with time in the *TV Corpus*, and the fact that they are found more frequently in the American portion of the corpus than in other subcorpora, is in line with studies pointing to the rise of familiarity in American English which, according to Murray (2002), is driven by teenagers and young adults.

Finally, our study shows that caution must be taken when performing quantitative analyses of the *TV Corpus*, especially because of the smaller number of tokens included in the first four decades in the corpus, and because these early years feature a higher proportion of documentaries, which seem to use quite a formal language, than later decades. We have to remember that, like all diachronic corpora, the *TV Corpus* is, despite its large size, a “narrow lens that provides an insight into the language that has been preserved” (Brezina 2018, 222). Conducting inferential analyses is thus not sufficient to generalize the findings to a whole language variety.

References

- Baños, Rocío, Silvia Bruti, and Serenella Zanotti. 2013. “Corpus Linguistics and Audiovisual Translation: In Search of an Integrated Approach.” *Perspectives* 21.4: 483–490. <https://doi.org/10.1080/0907676X.2013.831926>
- Baumgarten, Nicole. 2008. “Yeah, That’s It!: Verbal Reference to Visual Information in Film Texts and Film Translations.” *Meta* 53.1: 6–25. <https://doi.org/10.7202/017971ar>
- Bednarek, Monika. 2011. “The Language of Fictional Television: A Case Study of the ‘Dramedy’ *Gilmore Girls*.” *English Text Construction* 4.1: 54–84. <https://doi.org/10.1075/etc.4.1.04bed>
- Bednarek, Monika. 2012. “‘Get Us the Hell out of Here’: Key Words and Trigrams in Fictional Television Series.” *International Journal of Corpus Linguistics* 17.1: 35–63. <https://doi.org/10.1075/ijcl.17.1.02bed>
- Bednarek, Monika. 2015. “‘Wicked’ Women in Contemporary Pop Culture: ‘Bad’ Language and Gender in *Weeds*, *Nurse Jackie*, and *Saving Grace*.” *Text & Talk* 35.4: 431–451. <https://doi.org/10.1515/text-2015-0011>
- Bednarek, Monika. 2018. *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge: Cambridge University Press.
- Bednarek, Monika. 2023. *Language and Characterisation in Television Series: A Corpus-informed Approach to the Construction of Social Identity in the Media*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.
- “Bro” *Etymonline*. <https://www.etymonline.com/search?q=bro>

“Bro culture” *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=Bro_culture&oldid=1075514926

Bruti, Silvia, and Elisa Perego. 2010. “Vocatives in Subtitles: A Survey across Genres.” *Ecolingua: The Role of E-Corpora in Translation and Language Learning*. Ed. Christopher Taylor. Trieste: Edizioni Università di Trieste. 61–75.

Clancy, Steven J. 1999. “The Ascent of Guy.” *American Speech* 74.3: 282–297.

Csomay, Enico, and Ryan Young. 2021. “Language Use in Pop Culture over Three Decades: A Diachronic Keyword Analysis of Star Trek Dialogues.” *International Journal of Corpus Linguistics* 26.1: 71–94. <https://doi.org/doi:10.1075/ijcl.00037.cso>

Davies, Mark. 2021. “The TV and Movies Corpora: Design, Construction, and Use.” *International Journal of Corpus Linguistics* 26.1: 10–37.

Eitelmann, Matthias, and Dagmar Haumann. 2023. “Getting Close-ish: A Corpus-based Exploration of *-ish* as a Marker of Approximation and Vagueness.” *Zeitschrift Für Wortbildung/Journal of Word Formation* 7.1: 76–100. <https://doi.org/10.21248/zwjw.2023.1.91>

Forchini, Pierfranca. 2013. “A Diachronic Study of Familiarizers (‘man’, ‘guys’, ‘buddy’, ‘dude’) in Movie Language.” *Perspectives* 21.4: 504–525. <https://doi.org/10.1080/0907676X.2013.831923>

Formentelli, Maicol. 2014. “Vocatives Galore in Audiovisual Dialogue: Evidence from a Corpus of American and British Films.” *English Text Construction* 7.1: 53–83. <https://doi.org/10.1075/etc.7.1.03for>

Gregori-Signes, Carmen. 2017. “‘Apparently, Women Don’t Know How to Operate Doors’: A Corpus-based Analysis of Women Stereotypes in the TV Series *3rd Rock from the Sun*.” *International Journal of English Studies* 17.2: 21–43. <https://doi.org/10.6018/ijes/2017/2/257311>

Heyd, Theresa. 2010. “How You Guys Doin’? Staged Orality and Emerging Plural Address in the Television Series *Friends*.” *American Speech* 85.1: 33–66. <https://doi.org/10.1215/00031283-2010-002>

Hilbe, Joseph M. 2011. *Negative Binomial Regression*. Cambridge: Cambridge University Press.

Hill, Richard A. 1994. “You’ve Come a Long Way, Dude: A History.” *American Speech* 69.3: 321–327. <https://doi.org/10.2307/455525>

Hirota, Tomoharu, and Laurel J. Brinton. 2023. “‘You Betcha I’m a Merican’: The Rise of YOU BET as a Pragmatic Marker.” *International Journal of Corpus Linguistics*. <https://doi.org/10.1075/ijcl.21060.hir>

Johnson, Adrienne Rose. 2012. “Romancing the Dude Ranch, 1926–1947.” *Western Historical Quarterly* 43.4: 437–461.

Jones, Joseph R. 2017. “Brojobs, Bro Culture, and Hegemonic Masculinity: A Call to Reconceptualize How Educators Address Homophobia.” *Feather Boas, Black Hoodies, and John Deere Hats. Discussions of Diversity in K-12 and Higher Education*. Ed. Joseph R. Jones. Rotterdam, Boston & Taipei: Brill. 101–103.

- Jucker, Andreas H. 2021. “Features of Orality in the Language of Fiction: A Corpus-based Investigation.” *Language and Literature* 30.4: 341–360. <https://doi.org/10.1177/096394702111047751>
- Jucker, Andreas H., and Daniela Landert. 2023. “The Diachrony of Im/politeness in American and British Movies (1930–2019).” *Journal of Pragmatics* 20.9: 123–141. <https://doi.org/10.1016/j.pragma.2023.02.020>
- Kang, Namkil. 2022. “On Speak to and Talk to: A Corpora-based Analysis”. *Theory and Practice in Language Studies* 12.7: 1262–1270. <https://doi.org/10.17507/tpls.1207.03>
- Kiesling, Scott F. 2004. “Dude.” *American Speech* 79.3: 281–305.
- Kleinknecht, Friederike. 2013. “Mexican Güey – from Vocative to Discourse Marker: A Case of Grammaticalization?” *Vocative!: Addressing between System and Performance*. Ed. Barbara Sonnenhauser, and Patrizia Noel Aziz Hanna. Berlin: Walter de Gruyter. 235–268.
- Landert, Daniela, Tanja Säily, and Mika Hämäläinen. 2023. “TV Series as Disseminators of Emerging Vocabulary: Non-codified Expressions in the TV Corpus.” *ICAME Journal* 47.1: 63–79. <https://doi.org/10.2478/icame-2023-0004>
- Leech, Geoffrey. 1999. “The Distribution and Function of Vocatives in American and British English Conversation.” *Out of Corpora: Studies in Honour of Stig Johansson*. Ed. Hilde Hasselgård, and Signe Oksefjell. Amsterdam & Atlanta: Rodopi. 107–120.
- Levshina, Natalia. 2017. “Online Film Subtitles as a Corpus: An n-gram Approach”. *Corpora* 12.3: 311–338. <https://doi.org/10.3366/cor.2017.0123>
- Li, Hanxi, Honggang Liu, and Dilin Liu. 2022. “Gender/Power Relationships in Fictional Conflict Talk at the Workplace: Analyzing Television Dramatic Dialogue in *The Newsroom*”. *Journal of Pragmatics* 18.7: 58–71. <https://doi.org/10.1016/j.pragma.2021.10.030>
- Lu, Siwen. 2023. “How Existing Literary Translation Fits into Film Adaptations: The Subtitling of Neologisms in Harry Potter from a Multimodal Perspective.” *Visual Communication* 0:0: 14703572221141960. <https://doi.org/10.1177/14703572221141959>
- Murray, Thomas E. 2002. “A New Look at Address in American English: The Rules Have Changed.” *Names—a Journal of Onomastics* 50.1: 43–61. <https://doi.org/10.1179/nam.2002.50.1.43>
- Pastorino, Valeria. 2022. “Dude in British English: Towards a Non-gendered Term of Address.” *York Papers in Linguistics* 2.17: 13–28.
- Quaglio, Paulo. 2008. “Television Dialogue and Natural Conversation.” *Corpora and Discourse: The Challenges of Different Settings*. Ed. Annelie Ädel, and Randi Reppen. Amsterdam & Philadelphia: John Benjamin Publishing Company. 189–210

- Quaglio, Paulo. 2009. *Television Dialogue: The Sitcom Friends vs. Natural conversation*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- R Core Team. 2018. “R: A Language and Environment for Statistical Computing.” [Manual]. <https://www.R-project.org/>
- Reichelt, Susan. 2018. “The Sociolinguistic Construction of Character Diversity in Fictional Television Series.” PhD diss, Cardiff University.
- Rodríguez-Abruñeiras, Paula. 2022. “‘Me Likey!’ A New (old) Argument Structure or a Partially Fixed Expression with the Verb Like?” *Círculo de Lingüística Aplicada a La Comunicación* 90 :237–249. <https://doi.org/10.5209/clac.77163>
- Schwiegershausen, Erica. 2013. *Exploring the Etymology of ‘Bro.’* The Cut. <https://www.thecut.com/2013/10/exploring-the-etymology-of-bro.html>
- Sonnenhauser, Barbara, and Patrizia Noel Aziz Hanna. 2013. “Introduction: Vocative!*”. *Vocative! Addressing between System and Performance*. Ed. Barbara Sonnenhauser, and Patrizia Noel Aziz Hanna. Berlin: Walter de Gruyter. 1–24.
- Tagliamonte, Sali A. 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>
- The TV Corpus*. n.d. *English-Corpora*. <https://www.english-corpora.org/tv/>
- Urichuk, Matthew, and Verónica Loureiro-Rodríguez. 2019. “Brocatives. Self-Reported Use of Masculine Nominal Vocatives in Manitoba (Canada).” *It’s Not All about You: New Perspectives on Address Research*. Ed. Bettina Kluge, and María Irene Moyna. Amsterdam & Philadelphia: John Benjamins Publishing Company. 355–372.
- Werner, Valentin. 2021. “A Diachronic Perspective on Telecinematic Language.” *International Journal of Corpus Linguistics* 26.1: 38–70.