To what extent can a comedy drama provide a classroom model for natural conversation?

A multi-dimensional register analysis of the British series *Cold Feet*

Jill Cannon-Jones & Viola Wiegand
University of Stirling

The aim of this study is to determine whether the British comedy drama *Cold Feet* could be used in English learning contexts to expose students to useful elements of natural conversation. Using additive multi-dimensional register analysis, we empirically ascertain how similar the language used in *Cold Feet* is to that of natural conversation. Through *n*-gram analysis, we furthermore examine to what extent the subtitles of *Cold Feet* share any register markers or style features with natural conversation in the Spoken BNC2014. The results show that *Cold Feet* shares many characteristics with natural conversation. It is matched closest to "conversations" on Dimension 1, but "general fiction" on Dimensions 3 and 5. Based on the findings, we provide concrete pedagogical recommendations for using the comedy drama as a classroom model with which students can critically engage to improve their linguistic awareness and conversation skills.

Keywords: language learning, multi-dimensional analysis, register analysis, TV shows, telecinematic language

1. Introduction

Many pedagogical guides for teachers of foreign languages recommend using clips from television shows and films to help students improve their listening skills (e.g. Wilson 2008) or pragmatic skills (Washburn 2001). While this exposure is likely to improve students' understanding of similar television shows in the future, the purpose of this study is to examine whether, by watching clips of the British com-

https://doi.org/10.1075/rs.24012.can | Published online: 28 July 2025
Register Studies 8:1 (2026), pp. 1–27. ISSN 2542-9477 | E-ISSN 2542-9485
Available under the CC BY 4.0 license. © John Benjamins Publishing Company

edy drama *Cold Feet*, students would be exposed to language comparable to naturally occurring conversation.

This study aims to answer the following research questions:

1. To what extent does the language of *Cold Feet* resemble that of natural conversation?
2. How can extracts from *Cold Feet* inform activities in English language lessons?

The study will use an additive multi-dimensional register analysis combined with *n*-gram analysis to try to empirically answer these questions. It will then be possible to determine whether the use of *Cold Feet* in an English learning classroom has the potential to not only help the learners understand television programmes but, if it resembles natural conversation closely enough, to also offer the learners practice of understanding spoken language in a real-world situation and provide a model of natural conversation with which students can critically engage. The following sections will review previous research done in this area (Section 2) and introduce the data collection and analysis methods used in the present study (Section 3). While Section 4 presents the results from the corpus linguistic analyses, we provide concrete pedagogical recommendations for using *Cold Feet* as a classroom model for speaking activities in Section 5. Finally, Section 6 concludes the paper with suggestions for further research.

2. Telecinematic language and language teaching

In Section 2.1, we introduce relevant previous research on telecinematic language, focusing on register studies and comparisons with natural conversation. Section 2.2 then summarises the literature on using telecinematic fictional speech as a classroom model.

2.1 Telecinematic language and comparisons with natural conversation

Numerous previous studies have examined the language of television programmes and films, also known as 'telecinematic language' (Piazza, Bednarek, & Rossi 2011). Typically, these studies tend to focus either on film (e.g. Forchini 2012, 2021; Kozinski 2011) or on television programmes (e.g. Quaglio 2009; Al-Surmi 2012; Bednarek 2018). Exceptions include Werner's (2021a) study using the two english-corpora.org TV and Movie corpora and Csomay and Petrović's (2012: 308) study of television programmes and films related to "legal content".

Studies comparing telecinematic language to natural conversation have tended to use a register analysis framework (e.g. Quaglio 2009; Al-Surmi 2012; Forchini 2012; Veirano Pinto 2014). Biber and Conrad (2009: 31) define 'register' as "a language variety associated with both a particular situation of use and with pervasive linguistic features that serve important functions within that situation of use". They explain that whilst most linguistic features, such as lexical items and grammatical structures, could appear in a variety of text types, the core linguistic features for a particular register will be much more common in that register than in other text types. If the language in *Cold Feet* is similar to natural conversation, it should contain a similar pattern of linguistic features. According to Biber (2023), there are many advantages in using a corpus linguistic approach for register analysis; most importantly, the ability to use a computer to analyse a large group of texts rather than being restricted to analysing a few samples manually. Similarly, Biber and Conrad (2009) argue that it is not possible to distinguish among registers based on the frequency of just a few linguistic features: it is important to consider how these features co-occur. Multi-dimensional (MD) analysis examines these groups of linguistic co-occurrences known as 'dimensions': if various dimensions are considered and compared, it is possible to ascertain how and to what extent various registers differ from one another (Biber & Conrad 2009).

Corpus studies have identified a range of overlaps between telecinematic and natural conversation. For example, Forchini's (2012: 119) MD analysis of US film dialogues and face-to-face conversation in the Longman Spoken American Corpus (LSAC) showed that these domains "do not differ to a great extent" as both are "informal, non-narrative, situation-dependent and not highly persuasive". In particular, the film transcript scores showed the same polarity as the face-to-face data for four out of Biber's (1988) five dimensions (Forchini 2012: 90–91). An MD analysis comparing the sitcom *Friends* to the American conversation subcorpus of the Longman Grammar Corpus indicated that the linguistic features of *Friends* were, overall, very similar to those of natural conversation (Quaglio 2009). This similarity was evident through the "core linguistic features that characterize involved registers" (Quaglio 2009: 148) found in the sitcom on Dimension 1 "Informational versus Involved Production", including, e.g. first and second person pronouns, contractions, and present tense verbs (see Quaglio 2009: 68). Regarding situation-dependent features on Dimension 3, Al-Surmi's (2012) comparison of *Friends* with the soap opera *The Young and the Restless* found that both shows shared such features with natural conversation. Bednarek (2018) used a keyness comparison to identify words and *n*-grams (multi-word frequencies) in a large corpus of US American drama and comedy television transcripts compared with the LSAC. Among the similarities that her comparison identified were "features that reflect the characteristics of spoken interactions

(as non-monologic, spontaneous, on-line, expressive, informal, situation-bound talk)" (Bednarek 2018: 153).

Despite the large extent of overlap that has been found between telecinematic discourse and natural conversation, "[t]here is no doubt that in terms of literal spontaneity, movie language is fictitious: it pretends to be authentic, but it is planned and artificial" (Forchini 2012: 35). Accordingly, corpus evidence has also indicated some ways in which the domains do not entirely overlap, with MD analyses indicating that Dimension 5 "Abstract versus Non-abstract Information" occasionally differs. US American films (Forchini 2012), as well as the soap opera and sitcom studied by Al-Surmi (2012) scored positively on this Dimension, while natural conversation scored negatively in these studies and in Biber (1988). Forchini (2012: 91) describes the difference in scores as "fairly minimal", attributing it mainly to a "relatively higher presence of adverbial conjuncts" in the films. Al-Surmi (2012) found a higher discrepancy between Dimension 5 scores for the two television programmes and natural conversation, with the difference more pronounced for the soap opera *The Young and the Restless*. He explained the higher scores with passive constructions in the sitcom *Friends* and linked the soap opera to "elaborated discussions about business issues, court cases, and crimes" (see Al-Surmi 2012: 689–690). Additionally, Quaglio (2009) reported that *Friends* contained fewer examples of vague and narrative language but more informal and emotional language than natural conversation.

Corpus studies on telecinematic discourse have worked with a range of data sources (Veirano Pinto 2018). Forchini (2012, 2021: 190) emphasises the importance of working with manually created "authentic transcriptions of movie dialogs" as the most accurate data source for linguistic analysis. At the same time, Veirano Pinto (2018) shows that MD analyses of faithful transcripts and subtitles from the same films indicate similar results. Although subtitles tend not to include all spoken features such as pauses or interjections, these are not usually part of the variables of "mainstream" MD analysis (Veirano Pinto 2018: 150). In the present study, we work with online subtitles because these are more accessible for teachers who might want to apply our recommendations for practice (see Section 5) and are unlikely to have the time and resources to transcribe episodes from scratch.

While much previous research has tended to focus on North American television series and films (e.g. Quaglio 2009; Al-Surmi 2012; Forchini 2012), the present study focuses specifically on a British English television show. There have been some previous studies looking at the language in British television shows and films (e.g. Kozinski 2011; Lazzeretti 2019; Jones 2017). However, we are not aware of a multi-dimensional register analysis conducted on a British television show, especially with a pedagogical focus.

2.2 Telecinematic fictional speech as a classroom model for English language speaking skills

Research on the use of videos in general has attracted increasing attention in second language acquisition and language learning/teaching research; Montero Perez (2022: 163) notes "an exponential increase in studies on the effects of video on L2 learning" in the previous decade. This body of research has considered a wide range of aspects, including more technological ones such as the use of target-language captions in comparison with L1 subtitles, the relationship between video input and learning gains in relation to e.g. vocabulary, grammatical knowledge, and listening comprehension, as well as learners' perceptions (Montero Perez 2022). At the same time, the use of telecinematic dialogue in teaching relates to an increasing research interest in the use of pop culture materials in language teaching especially because of "their potential to affectively engage learners" (Werner & Tegge 2020: 7). Despite the popularity of US-centric television series and films (see previous section), pop culture materials can be selected with a view to representing a diversity of speakers. For example, Bagni and Sumra (2022) propose ways of using pop culture materials informed by the Global English Language Teaching paradigm that promotes principles such as raising awareness of and promoting respect for World Englishes and English as a Lingua Franca.

Previous studies have suggested using extracts from television series or films as the basis for classroom activities. In a study of the British soap opera *Eastenders*, Jones (2017) argued that because soap operas are more easily accessed by teachers than unscripted conversations, they can be a good source of material, especially for raising students' awareness of conversation. Al-Surmi (2022: 50) suggests that extracts from the sitcom *Friends* and the soap opera *The Young and the Restless* can serve as different types of input in class, arguing e.g. that the sitcom provides a better model of disfluency and vagueness, while the soap opera better illustrates the use of stance expressions. When working with extracts, teachers can highlight common linguistic features in bold as a basis for discussing their functions in spoken language, as Forchini (2021) illustrates with film dialogues.

Regarding the positive potential of extracts from telecinematic discourse, a key consideration is how the material is introduced and how the learners are encouraged to engage with it. The design of activities needs to ensure that telecinematic "corpus data is not used as a model for learners to mimic slavishly but as examples of communicative behaviour against which they can critically compare their own performances" (Anderson & Corbett 2009: 177). Critical engagement with extracts from television series has the potential to facilitate, for example, reflective debates about stereotypes in the representation of social groups, and thereby contribute to developing both students' linguistic awareness and 'media literacy' (Bednarek

2018: 246). With the results from the present study and the practical recommendations in Section 5, we aim to build on and contribute to the existing work on the use of telecinematic discourse as a model in English language teaching.

3. Methodology

Section 3.1 introduces the television show *Cold Feet* and our corpus compilation procedure. In Section 3.2, we introduce the steps taken for the MD analysis and the extraction and analysis of *n*-grams.

3.1 The *Cold Feet* corpus

Cold Feet is a British comedy drama set in Manchester and follows the lives of a group of (adult) friends (Rotten Tomatoes 2023). It was chosen for similar reasons to those given by Quaglio (2009) for choosing *Friends*: both shows were very popular, with *Cold Feet* running for nine series (Rotten Tomatoes 2023), and depict scenes from everyday life. Quaglio (2009: 13) chose *Friends* to compare to natural conversation because it was "a show about people who just sit around and talk". Similarly, *Cold Feet* was described as "a portrait of urban life as it really is" (Barber 1998) which suggests that it may contain realistic dialogues suitable for exposing students to natural sounding conversation.

The episodes of *Cold Feet* analysed in the present study are taken from the following series, for which we were able to retrieve subtitles from Subslikescript (2024): 5th (2003), 6th (2016), 8th (2019), and 9th (2020). In order to check the accuracy of the subtitles from Subslikescript (2024), the subtitles for the most recent available episode, series nine episode 4 (S9E4), were carefully compared to the corresponding video. Only three minor errors were found in the whole episode, for example "Jen" was transcribed once as "Jane". Such a low number of errors shows a very high accuracy rate of the subtitles. However, there were some repetitions, false starts, or fillers (e.g. *um*, *er*) that had not been included. The original transcripts were provided without character names or any major additional "noise". To reduce any issues falsely affecting the tagging of the dialogue for the MD analysis (see Section 3.2.1), we removed text not belonging to the subtitles (advertising, subtitle and soundtrack credits), hyphens at beginning of lines, non-verbal cues in all capitals (only present in one episode), and soundtrack lyrics (that were indicated by the symbols # or ♪; we checked these scenes and only kept lyrics sung by characters).

The ten subtitle files were uploaded to Sketch Engine (Kilgarriff, Baisa, Bušta, Jakubíček, Kovář, Michelfeit, Rychlý, & Suchomel 2014; http://www.sketchengine

.eu, accessed in July 2024) to create a corpus. The resulting *Cold Feet* corpus contains 67,666 tokens (a figure which includes words, punctuation and numbers), 50,095 of which are words. Table 1 shows an overview of the final corpus and the composition of episodes.

Table 1. *Cold Feet* corpus size and composition (following the series/episode numbering by Subslikescript 2024)

Series (S) and episode (E) number	Number of tokens
S5E1	7,177
S5E2	6,377
S5E3	6,733
S5E4	6,689
S6E5	7,312
S8E1	7,405
S8E2	7,043
S8E3	5,728
S8E4	6,969
S9E4	6,233
Total	67,666

3.2 Analysis methods

This section first introduces the procedures followed for the additive MD analysis using MAT (3.2.1). Section 3.2.2 then reports on the steps taken for the *n*-gram analysis using Sketch Engine.

3.2.1 *Additive MD analysis*

The first step for the study was to outline the situational characteristics for the *Cold Feet* corpus. An additive MD analysis was used to compare the MD analysis results from a new register relative to other registers that have previously been identified (Berber Sardinha 2014). In the present study, the results from the MD analysis were compared to Biber's (1988) dimension scores to determine how closely the *Cold Feet* corpus resembles natural conversation.

To carry out the additive MD analysis, this study used the Multidimensional Analysis Tagger ("MAT"; Nini 2022). Although many studies conduct MD analyses using the Biber Tagger (including e.g. Quaglio 2009; Al-Surmi 2012; Forchini 2012), this program is not publicly available. The original Biber tagger, developed by Biber (1988), tags 67 linguistic features in a corpus. MAT is a free program which "replicates Biber's (1988) tagger" (Nini 2019: 71): it generates dimension scores directly comparable with Biber (1988), produces plots for each dimension, and matches the texts with Biber's (1988) genres and Biber's (1989) eight text types. The software has since been used for additive MD analyses of, for example, textbook registers (Le Foll 2021) and pop lyrics Werner (2021b), and as part of a method to identify informal YouTube videos (Cooper 2023).

The plain text files were uploaded to MAT which first uses the Stanford Tagger (Toutanova, Klein, Manning, & Singer 2003) to tag basic parts of speech (nouns, verbs etc.) and then tags more complex grammatical features that were also identified with the Biber tagger, e.g. *that* used as a demonstrative adjective (Nini 2019). Based on the tagged files, MAT calculated dimension scores for the corpus. It initially produced two files: one containing the relative frequencies of the linguistic features per hundred words and the other showing the "standardized scores, or z-scores" (Nini 2019: 71). MAT added together the z-scores for all the positive features of a dimension and subtracted the z-scores for the features not characteristic of a particular dimension to give the overall dimension score (Nini 2019). Finally, MAT generated a graph for each of Biber's (1988) six dimensions showing the mean and range for the dimension of each type of text including the corpus currently being studied.

In order to replicate the Biber (1988) tagger as closely as possible and enable an effective comparison with previous results based on Biber's (1988) dimensions and Biber's (1989) text types, we used the recommended MAT settings (z-score correction: no, count: only VASW tags and 400 tokens for type/token ratio). The z-score correction can be used on MAT to eliminate any z-scores higher than 5 (or lower than -5) which can help to remove any unlikely values which could skew the results (Nini 2019). The VASW tags are the tags used in Biber's (1988) *Variation across speech and writing*, which, in that original study, were applied to the first 400 tokens of each text (see Nini 2020). We tested whether altering the number of tokens checked per text would have an effect on the results but found that this was not the case (when we ran MAT on the *Cold Feet* corpus with a setting of 5,000 tokens, the closest text types and genres assigned to the corpus remained the same). To facilitate compatibility with the previous Biber Tagger results, the findings reported in Section 4 are therefore based on the recommended MAT settings (i.e. covering just the first 400 tokens).

The present study focuses on the three dimensions that Biber (1988: 160–162) found to be the most relevant to conversation. Taking natural conversation and academic writing as the two extremes of an oral/literate dimension, Biber (1988) found that for Dimensions 1, 3, and 5, these two types of texts fell at opposite ends of the scale. This contrasts with Dimensions 2, 4, and 6, where they were much more closely grouped (Biber 1988). As Biber (1988), therefore, considered Dimensions 1, 3, and 5 as oral/literate dimensions, we decided to focus on these three in the present study (see also Werner 2021b). More information about the nature of each dimension is given in Section 4.2.

3.2.2 N-*gram analysis*

N-grams are groups of two or more words that occur together repeatedly in a corpus, which research has shown will vary depending on the register of that text (e.g. Biber, Conrad, & Cortes 2004; Cortes 2015). We extracted *n*-grams using Sketch Engine in order to identify linguistic features that are shared with natural conversation. The decision to complement MD analysis with *n*-gram analysis is inspired by Werner (2021b), who examined *n*-grams in his study of pop lyrics to identify linguistic features that might have been missed by the MD analysis on its own. For our aim of making pedagogical recommendations, the focus on *n*-grams has the additional advantage that it can help with developing activities and teaching materials based on formulaic language rather than single words. Vu and Peters (2022: 288), for example, highlight the "ubiquity of FSs [formulaic sequences] in speech, the role of FSs in oral proficiency (especially oral fluency), and the challenges of learning FSs", which therefore require special attention in class, including through authentic input e.g. from television, explicit instruction, and opportunities to produce FSs (see Vu & Peters 2022: 292–293).

For generating *n*-grams, the minimum frequency was set to "3" to eliminate any very rare words. We chose the Spoken BNC2014 (Love, Dembry, Hardie, Brezina, & McEnery 2017), available from Sketch Engine, as a suitable reference corpus because both the time frame and the type of participants overlap with the *Cold Feet* series analysed here. The Spoken BNC2014 consists of informal conversations between family and friends recorded between 2012 and 2016, features 668 British English speakers, and contains 11.5 million words (Love et al. 2017). Supplementing the MD analysis with this comparison to the Spoken BNC2014 allowed us to balance out the potential limitation that the conversational data underlying Biber's (1988) dimensions in our MD analysis stems from older decades, the 1960s–1980s (see also Cooper 2023). We generated case-insensitive *n*-grams (excluding nonwords) of a length of three words ('3-grams') due to the relatively small size of the corpus (containing a limited number of longer *n*-grams) and then manually compared the 50 most frequent 3-grams displayed

by Sketch Engine (spread across 19 frequency ranks in the *Cold Feet* corpus, due to some shared frequencies) with the 50 most frequent 3-grams in the Spoken BNC2014 corpus (spread across 49 frequency ranks) to identify the extent of the overlap. The cut-off was chosen due to space constraints. Through a mixture of concordance analysis and comparing the *n*-gram list with the literature on the linguistic features of natural conversation (Biber & Conrad 2009; Biber 2023; Cortes 2015), we examined potential similarities and differences.

4. Results

This section first explores the situational characteristics of the *Cold Feet* corpus (4.1). Subsection 4.2 then presents the results from MAT. The findings of the *n*-gram analysis are discussed in 4.3.

4.1 Situational characteristics

At first glance, the situational characteristics for *Cold Feet* and natural conversation are very similar. Quaglio (2009) identifies the following features of natural conversation: highly interactive, face-to-face, real-time communication, in shared and mostly informal contexts. These are features which *Cold Feet* appears to share. The conversations in the show involve the five main characters, who are friends, talking to each other or to other people in their lives, such as family members and work colleagues. Their conversations take place in a wide range of settings, which in S9E4 include scenes in the characters' houses, the hospital, their offices, a cafe, at the gym, and at a funeral service. Conversations in the drama generally take place face-to-face, with the exception of some telephone calls, and in real-time.

These situational characteristics arguably give rise to the "traces of spontaneity" evident from telecinematic language, i.e. overlap in linguistic features with natural conversation (Forchini, 2012: 36). As Gregory and Carroll (1978: 42) argue, "[w]hen the actor performs, having learnt his [or her] lines and rehearsed them, he [or she] is speaking what is written to be spoken as if not written". While defining the situational characteristics of a piece of fiction can be complicated, "[t]he determining factors for the language used in a fictional text are the ways in which the fictional world is constructed, rather than the 'real-world' situation of the text" (Biber & Conrad 2009: 132). What matters are how the "audience perceives" the situational characteristics (see Forchini 2012: 37). In other words, although *Cold Feet* is based on a written script, because the fictional world created is one where the characters have face-to-face interactions, we could predict that the language

used will resemble natural conversation. In the present study, we aim to determine to what extent this is the case for the *Cold Feet* corpus.

As *Cold Feet* is a television show, "the telling of a narrative, one which might absorb, entertain, inspire, and move the viewer" (Toolan 2011: 181) can be considered its main communicative purpose. It therefore seems likely that there will be a higher number of dramatic interactions such as arguments and conflict and fewer mundane everyday conversations such as discussing what to watch on television or when the kids need picking up. As Washburn (2001: 22) notes, "television characters seem to have much more eventful lives than the rest of us". This could explain why Quaglio (2009) found more emotional language in *Friends* than in natural conversation. The reference corpus used in the present study (Spoken BNC2014) contains conversations recorded with the participants' consent. As the compilers point out, it is not possible for the corpus to be a statistically accurate representation of the whole language (Love et al. 2017). There are "selected characterisations of conversation type" listed in the corpus, such as "discussing", "explaining", "inviting", "telling jokes" etc. but apart from potentially "complaining" and "apologizing", with little indication of direct conflict (see the Sketch Engine 'Text type analysis' feature). Topics of conversation in the comedy drama *Cold Feet* may therefore often differ from those that people recorded for the Spoken BNC2014 corpus.

Indeed, there are some very common features of natural conversation that are almost always absent in television dramas: when speakers interrupt or overlap each other (Quaglio 2009). In *Cold Feet*, when S9E4 was compared in detail to the subtitles, only one occurrence of overlap was noted, indicated with square brackets in Example (1).

(1) **Karen:** *If I'd known you weren't going to uni, I'd have never [allowed it.]*
Liv: *[Mum,] I hadn't thought about it from your point of view.*

It seems likely that the lack of these features is due to a difference in situational characteristics. Although *Cold Feet* is trying to emulate natural conversation, this must be selective, as part of the wider aim of telling an (entertaining) narrative, requiring "coherent and focussed" dialogue (see Toolan 2011: 181). If telecinematic language contained as many overlaps and interruptions as natural conversation, it would be much more difficult for the audience to understand (Quaglio 2009).

4.2 MAT results

Firstly, MAT compares the target files to Biber's (1989) characteristic text types and finds the closest match (Nini 2019). MAT results show that for the whole *Cold Feet* corpus, as well as each of the individual episodes, the closest text type overall

is "informational interaction", a "[t]ext type that usually includes personal spoken interactions focused on informational concerns", with "face-to-face interactions, telephone conversations, spontaneous speeches, personal letters" being 'characterizing registers' (Nini 2019: 69). Like Werner (2021b), whose pop lyric corpus was assigned the same closest text type overall, we take this text type allocation as supporting evidence of similarity between the *Cold Feet* corpus and natural conversation.

MAT then calculates a score for each dimension. Table 2 shows the mean scores for Dimensions 1, 3, and 5 as the three most relevant dimensions for determining where a particular register falls on the speech-writing continuum (see Section 3.2.1).

Table 2. Mean dimension scores for the *Cold Feet* corpus

Dimension	Mean score
1	31.59
3	−1.95
5	−2.22

Dimension 1 is also called "Informational versus Involved Production" (Biber 1988: 115). A high score in Dimension 1 indicates that the text contains language associated with feelings and attitudes and involves direct interpersonal interaction (Biber 1988). Typical features of an involved discourse include (amongst many others) first and second pronouns, present tenses, possibility modals, indefinite pronouns, and private verbs such as *think* and *feel* (see Biber 1988: 105). Face-to-face conversations score highly on this dimension (mean score of 35.3) and academic, written texts have low scores (mean score of -14.9) (Biber 1988: 123–125). *Cold Feet* scored highly in this dimension with a mean score of 31.59 (see Figure 1). Furthermore, in the MAT comparison with 8 of the 'genres' examined in the original Biber (1988) study (face-to-face conversations, broadcasts, prepared speeches, personal letters, general fiction, press reportage, academic prose, and official documents; see Nini 2020), MAT indicated that, for this dimension, the closest genre to *Cold Feet* was "conversations" or 'face-to-face conversations' following Biber's (1988: 125) terminology. Many examples of the common features of this dimension can be found in the corpus as illustrated in (2), (3), and (4).

(2) ***I'm** not sure **I know** what **you mean**.* (first person pronouns, contractions, second person pronouns, private verbs) [S5E3]

(3) *If you **wanna** be a doormat, it's up to you but he's not **taking** advantage of me.* (present tenses) [S5E4]

(4) *Well, she **might** be able to hear you.* (possibility modals) [S9E4]

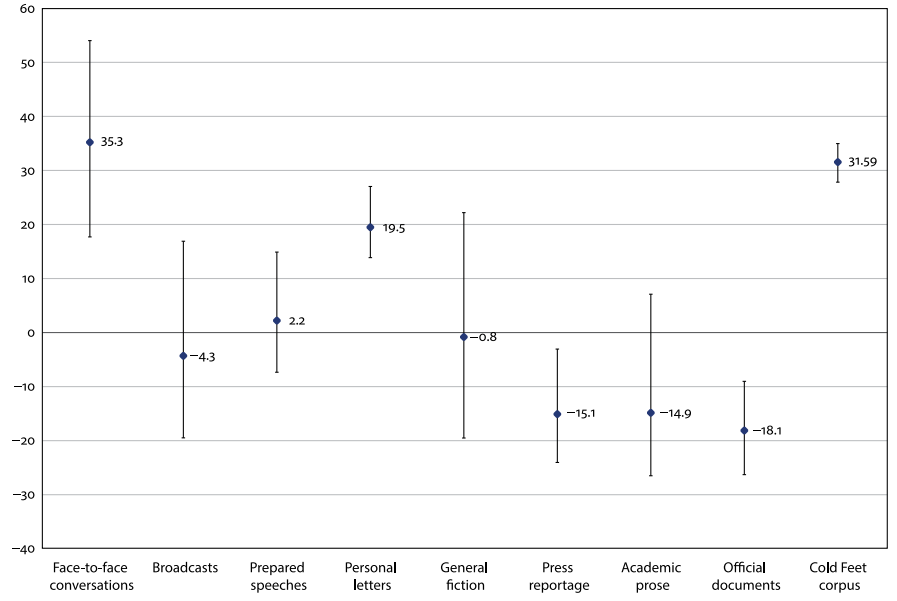


Figure 1. MAT Dimension 1 scores for the *Cold Feet* corpus compared with scores for eight of Biber's (1988: 122–125) genres

While the Dimension 1 score for *Cold Feet* is slightly lower than Quaglio's (2009: 65) score for the US American sitcom *Friends* (34.4), the result compares favourably to the British natural conversation data in Le Foll (2021), who also used MAT to obtain a Dimension 1 score for the Spoken BNC2014 (see Section 3.2.2 for more details about this corpus). Le Foll (2021: 225) reports that MAT gave the Spoken BNC2014 a score of 26.02 for Dimension 1, which was lower than the *Cold Feet* score of 31.59. However, as MAT requires punctuation to identify five of the 22 features with positive loading in Dimension 1 and the Spoken BNC2014 does not include all punctuation marks, this score is artificially low (Le Foll 2021). Taking into account only the features that do not require punctuation for identification, Le Foll (2021: 227) estimated the actual Dimension 1 score to be 30.66, which is extremely close to the *Cold Feet* score.

For Dimension 3, also labelled "Explicit versus Situation-Dependent Reference" (Biber 1988: 115), a low score would indicate discourse that is dependent on its immediate context, such as a live sports broadcast (Biber 1988). Typical features of a situation-dependent discourse include time and place adverbials — illustrated with *Cold Feet* examples in (5) and (6) — and general adverbs. Face-to-face conversation has a fairly low score on this dimension (-3.9) but not as low

as broadcasts (-9.0) (Biber 1988: 125). *Cold Feet* received a moderately low score (-1.95), which was higher than the mean score for conversation but still within the range shown in Figure 2. The corpus was assigned "general fiction" as the closest genre within this dimension.

(5) *Oh, stay **tonight**, will you?* (time adverbial) [S9E4]

(6) ***Near** the shops.* (place adverbial) [S5E4]

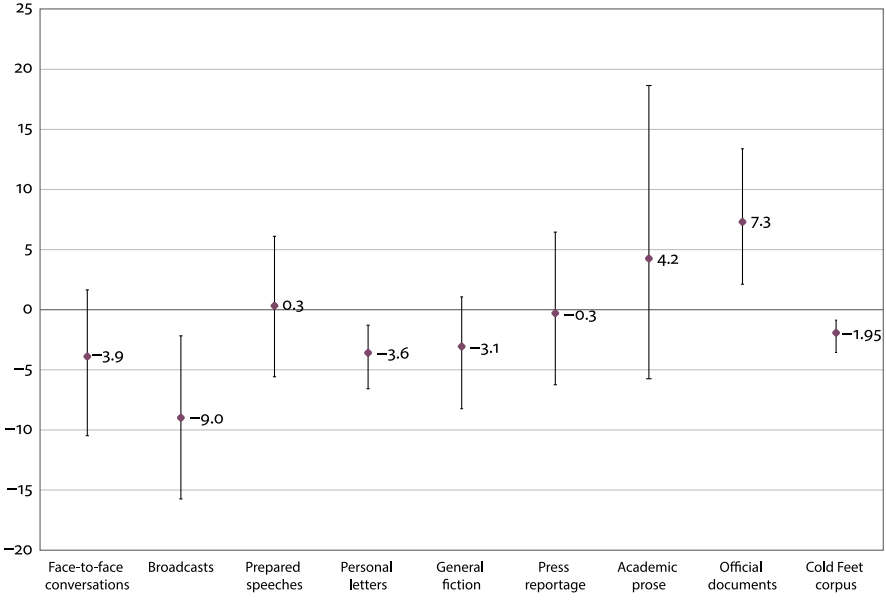


Figure 2. MAT Dimension 3 scores for the *Cold Feet* corpus compared with scores for eight of Biber's (1988: 122–125) genres

These Dimension 3 results were higher than the results Al-Surmi (2012: 687) reported for his sitcom and soap opera corpora: the mean score in this dimension for the *Friends* sitcom was -5.27 and for *The Young and the Restless* soap opera it was -5.62. Both of Al-Surmi's (2012) scores were lower than Biber's (1988) score for face-to-face conversation, which was -3.9.

Finally, in Dimension 5, which is labelled "Abstract versus Non-Abstract Information" (Biber 1988: 122), a high score indicates a high level of technical and abstract information (Biber 1988). Typical features of abstract information include conjuncts, agentless passives, and adverbial past participle clauses (Biber 1988). We would expect to find an absence of these abstract features in natural conversation, and indeed face-to-face conversations had a mean score of -3.2 in the original Biber (1988) study.

Similar to Dimension 3, in Dimension 5 *Cold Feet* scored moderately low (-2.22) but higher than Biber's (1988) score for face-to-face conversations. This finding, again, places the *Cold Feet* corpus closest to "general fiction" for Dimension 5 (see Figure 3). As the bars in Figure 3 indicate, the *Cold Feet* results fall within the conversation scores but do not cover their full range. This finding suggests that the range of language used in the television drama may be more limited than in real life conversations. Though focusing on Dimension 1 rather than Dimension 5, Quaglio (2009: 65) similarly noted that "*Friends* presents much less variation than conversation. This difference can be attributed to the wider range of situations and settings […], ages (from youngsters to octogenarians), and dialectal varieties which were 'captured' by the conversation corpus". As discussed in 4.1, moreover, some aspects of everyday conversations that are either too mundane or too difficult to follow (e.g. overlaps) would not be included in a drama designed to narrate and entertain.

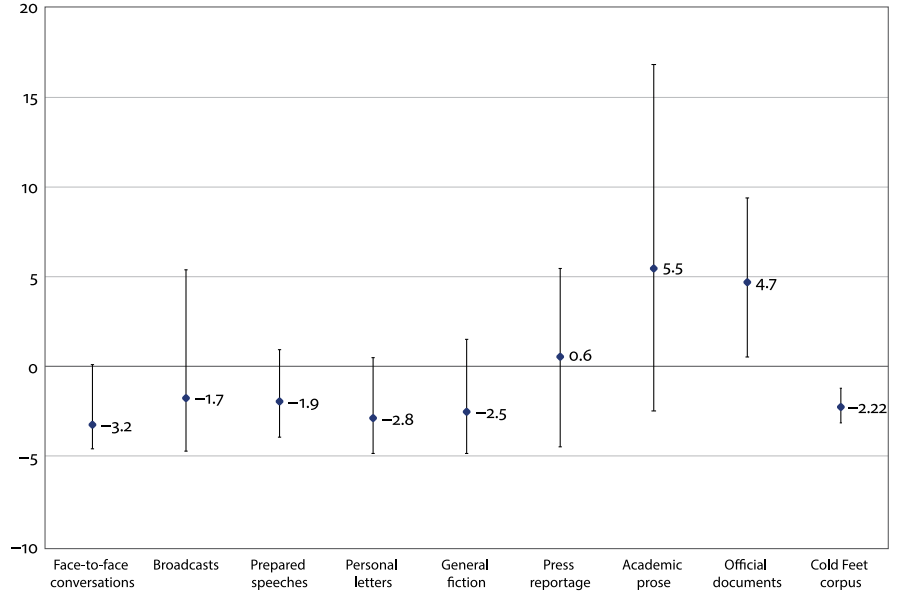


Figure 3. MAT Dimension 5 scores for the *Cold Feet* corpus compared with scores for eight of Biber's (1988: 122–125) genres

4.3 *N*-gram results

We now examine typical formulaic language in the *Cold Feet* corpus via *n*-grams. Table 3 shows the 50 most frequent 3-grams. Many of the features that have been identified by Biber (2023) as common in conversation can be seen in these

3-grams, including e.g. first or second person pronouns, contractions, and verbs expressing personal stance. The majority of the 3-grams (34 out of 50) contain a first or second person pronoun. Biber (2023: 309) explains that pronouns are "extremely frequent in conversation" because the participants often refer to themselves, each other, and the objects and people around them.

Table 3. The 50 most frequent 3-grams from *Cold Feet* corpus generated with Sketch Engine

Freq. rank	3-gram	Freq.	Freq. rank	3-gram	Freq.
1	i do n't	155	16	want you to	15
2	do n't know	53	16	you ca n't	15
2	you do n't	53	16	you have to	15
3	i ca n't	50	16	do n't need	15
4	i did n't	45	16	do you know	15
5	do n't want	35	17	i think i	14
6	do n't you	34	17	are you sure	14
7	n't want to	29	17	do n't worry	14
8	is n't it	28	18	did n't want	13
8	i want to	28	18	you did n't	13
8	what do you	28	18	i was n't	13
9	do n't think	27	18	i think it	13
9	do you think	27	18	n't have to	13
10	you want to	25	18	i love you	13
10	i thought you	25	18	why do n't	13
11	do you want	20	18	thought you were	13
12	i have n't	19	19	we do n't	12
12	what are you	19	19	to see you	12
13	do n't have	18	19	i have to	12
14	are you doing	17	19	i know you	12
14	a couple of	17	19	you wo n't	12
14	it does n't	17	19	there you go	12
15	have you got	16	19	i told you	12
15	you all right	16	19	are you all	12
15	going to be	16	19	to do with	12

The list of the 50 most frequent 3-grams includes 23 examples of contractions, ten of which are found across the top nine frequency ranks. Note that Sketch Engine counts the contraction *n't* as a separate word, leading to 3-grams such as *I do n't* in Table 4.[1] Contractions are associated with the situational context of real-time production where time pressure leads speakers to "take shortcuts with their language" (Biber & Conrad 2009: 90). Examples from *Cold Feet* are shown in (7) and (8).

(7) ***I don't** understand why we've got to come back.* [S5E3]

(8) *Mum, you're not telling us anything we **don't know**.* [S6E5]

According to concordance lines, 14 of the 3-grams occur (mostly or exclusively) as part of questions, e.g. *don't you*, *isn't it*, *what do you*, *do you think*. Further examples are given in (9) and (10). Biber and Conrad (2009) state that questions are associated with interactivity, which is a major feature of face-to-face conversation. Questions (typically) require a response and someone to give that response, so generally only occur in the context of some form of interaction (Biber & Conrad 2009).

(9) ***Are you all** right, mate?* [S8E2]

(10) *Why, **have you got** some more jobs that need doing?* [S9E4]

Additionally, the 3-grams include many private verbs expressing what Biber and Conrad (2009: 68) call 'personal stance' such as "personal pronoun + mental or desire verbs", as we can see in (11), (12), and (13). We also saw from the description of Dimension 1 that these are frequently found in natural conversation.

(11) *I **don't know** if it's jetlag or grief.* [S9E4]

(12) *I **don't want** your money!* [S6E5]

(13) *Mum, **I know you**'re upset about Jenny.* [S8E4]

1. *I'm* is another frequent contraction in the *Cold Feet* corpus (freq. = 457) but is not included in the *n*-grams, because Sketch Engine divides *I* and *'m*, considering *'m* as a "nonword". As the *n*-gram lists for *Cold Feet* and the Spoken BNC2014 "with nonwords" are dominated by sequences including punctuation, we focus on the more meaningful lists without "nonwords".

Table 4. The 50 most frequent 3-grams in the Spoken BNC2014 corpus generated with Sketch Engine

Freq. rank	3-gram	Freq.	Freq. rank	3-gram	Freq.
1	i do n't	35,567	26	and then i	3,016
2	do n't know	17,355	27	you know i	3,003
3	is n't it	9,298	28	do n't want	2,899
4	i ca n't	7,497	29	it does n't	2,891
5	you do n't	7,164	30	do you want	2,859
6	i did n't	6,867	31	a little bit	2,834
7	do n't think	6,498	32	do n't you	2,813
8	a lot of	6,455	33	yeah i think	2,741
9	i think it	6,070	34	i think that	2,701
10	i was like	5,776	35	it was like	2,588
11	yeah yeah yeah	5,545	36	oh my god	2,580
12	you ca n't	4,809	37	it was n't	2,544
13	i think i	4,241	38	and then you	2,537
14	and i was	3,971	39	n't know what	2,517
15	you have to	3,909	40	yeah yeah i	2,483
16	a bit of	3,681	41	yeah you know	2,424
17	i have n't	3,572	42	yeah i know	2,421
18	they do n't	3,529	43	you know it	2,404
19	no no no	3,490	44	yeah and then	2,389
20	and it was	3,353	45	do n't like	2,340
21	do n't have	3,245	46	one of the	2,339
22	do you know	3,133	47	you know you	2,292
23	you know what	3,087	48	be able to	2,290
24	i mean i	3,061	49	i know i	2,280
25	it was a	3,024	49	a couple of	2,280

There is considerable overlap between the 50 most frequent 3-grams found in the *Cold Feet* corpus and the Spoken BNC2014 (shown in Table 4). The two most frequent 3-grams, *I don't* and *don't know*, are the same in both corpora. Six of the first ten *Cold Feet* 3-grams listed in Table 3 are also in the top ten 3-grams in the Spoken BNC2014 corpus. Out of the four that are among the first ten listed for *Cold Feet*, but not in the Spoken BNC2014 corpus top ten, three include the word

want and form 'dependent clause fragments', reflecting the 'clausal' nature common in the phraseology of "spoken registers" (Cortes 2015: 207):

- *don't want*: frequency rank 5 in the *Cold Feet* corpus, 28 in the Spoken BNC2014
- *n't want to*: frequency rank 7 in the *Cold Feet* corpus, 67 in the Spoken BNC2014
- *I want to*: frequency rank 8 in the *Cold Feet* corpus, 122 in the Spoken BNC2014

This difference in the frequency ranks of 3-grams containing *want* could be due to the fact that the *Cold Feet* corpus is much smaller than the Spoken BNC2014 corpus, so it is not an ideal comparison. Another possible reason for the difference between the number of 3-grams containing the word *want* in the two corpora was hypothesised to be that in natural conversation participants might be more likely to say *wanna* instead of *want to*. A search for *wanna* in the *Cold Feet* corpus revealed only 7 hits which is equivalent to 103.45 occurrences per million tokens, compared to the Spoken BNC2014 where *wanna* had a raw frequency of 6,687 (565.12 occurrences per million tokens). A search for *want to* revealed a raw frequency of 98 (1,448.29 occurrences per million tokens) in the *Cold Feet* corpus and 7,588 occurrences (641.26 per million tokens) in the Spoken BNC2014 corpus. This shows that *want to* is more than twice as common in the *Cold Feet* corpus compared to the Spoken BNC2014, while the form *wanna* is much less prevalent than in the transcripts of natural conversation. Accordingly, the reason there are more 3-grams containing the word *want* among the top 3-grams in the *Cold Feet* corpus could relate to differing transcription protocols or differences in enunciation due to the different situational characteristics of the two corpora: actors often enunciate more clearly than in naturally occurring conversation to ensure the dialogue is easily understood by the audience (Bednarek 2018). There are eight occurrences of *want to* in the transcript for S9E4: a further rewatching of the episode showed that for two of the eight the actor actually said *wanna*. So, the difference between the number of 3-grams containing the word *want* could be due to a mixture of transcription and situational differences between the two corpora.

5. *Cold Feet* as a classroom model for natural conversation

Many practitioners and researchers have recommended that "authentic" material should be used in the classroom wherever possible (e.g. Basturkmen 2001; Thornbury & Slade 2006). However, it can be difficult for teachers to find record-

ings of natural conversations to use in the classroom and these can be overly complicated and difficult for learners to understand. Based on her MD analysis of English language textbooks used in Germany and France compared with the Spoken BNC2014, Le Foll (2021: 235) argues that "textbook dialogues appear to primarily function as reinforcers of the vocabulary students are expected to learn, rather than as models of realistic spontaneous interactions". Her results showed that textbook conversations only scored 15.75 on Biber's (1988) Dimension 1 (Le Foll 2021: 221) — compared to 35.3 for natural conversation in Biber's (1988) original MD analysis (see also Section 4.2). Apart from categories defined via punctuation, "the most underrepresented Dimension 1 features in Textbook Conversation are hedges, *that*-deletions, WH-clauses and *it* pronouns" (Le Foll 2021: 235).

Therefore, if authentic conversations are difficult to obtain and potentially too confusing for students, and textbooks do not always offer a reliable model of natural conversations, an alternative is needed. As we have shown here that the *Cold Feet* corpus contains many of the features found in natural conversation, we argue that extracts can be used as a classroom model for conversation. For example, if we want to expose students to the features from Dimension 1 that Le Foll (2021) found were lacking in textbooks, (14)–(17) show that there are numerous examples of them in *Cold Feet*.

(14) *I would give you my card but I'm* ***sort of*** *between jobs at the moment.* (hedge) [S8E3]

(15) *Well,* ***I suppose you*** *could pick out what you can afford.* (that-deletion) [S5E3]

(16) *Because I knew* ***what you'd think!*** (WH-clauses) [S6E5]

(17) ***I thought it*** *was you.* (that-deletion, pronoun *it*) [S9E4]

The following lesson outline builds on proposals from Washburn (2001) and Thornbury and Slade (2006) and shows how an extract from *Cold Feet* can be used in a lesson with the aim of introducing students to important features of natural conversation that are often missing from textbooks (Le Foll 2021). While these steps could be adapted for different groups and levels, we would suggest that the content of *Cold Feet* is most appropriate and interesting for upper intermediate to advanced adult learners (i.e. level B2 upwards). The CEFR descriptors state that B1 level students can understand "the main points in TV programmes on familiar topics when the delivery is relatively slow and clear" (Council of Europe 2020: 53); since that is not the case with *Cold Feet*, these students may struggle with comprehension of the dialogue. As *Cold Feet* is about the lives of five adults, adult students may find it easier to relate to the characters and consider the language useful to their own lives (Washburn 2001).

1. Students watch a clip of *Cold Feet* without sound and then discuss in groups what they think is happening in the scene.
2. Students watch the clip without sound again. This time, each member of the group takes one role in the scene and tries to voice the part.
3. Students watch the clip for the third time, now with sound, and make notes of any differences between their version of the scene and the original.
4. Students are given a handout to read with the transcript[2] for the scene. Depending on the focus of the lesson, they could be asked to underline or count a particular feature, complete a gap-fill exercise, or read it aloud in their groups.
5. Finally, students put the transcript away and watch the scene again, without sound, taking it in turns to voice the different roles.

Consider (18) as a possible example to be used in class. In this extract from S8E1, David has come to talk to Karen, his ex-wife, about financial arrangements. However, David then discovers that their son, Josh, is at home, and wants to drop out of university. The second turn — *I thought you'd be at university* — conveys David's surprise through the private verb *thought* and that-deletion. Other involved features include, for example, the discourse particle *well*, questions, first and second person pronouns, and contractions. This example also contains a filler, *erm*, and vague language (*that's the thing*), as Josh buys himself time to explain his decision not to finish his degree. In addition, the extract features two 3-grams from Section 4.3, *you don't* and *I thought you*, as examples of formulaic language to which the teacher could draw attention (see Section 3.2.2; Vu & Peters 2022).

(18) **David:** ***Josh?*** *Ah ...*

Josh: *Dad.*

David: ***I thought you'd*** *be at university.*

Karen: *Josh is just spending a few days at home ... to consider his future.*

David: *Oh,* ***you don't*** *graduate for another year.* ***What's to consider?***

Josh: ***W-Well, that's the thing. Erm ... I'm*** *not going to graduate.*

David: *Of course* ***you*** *are. Nowadays,* ***you*** *only have to turn up to the exams, they give* ***you*** *a 2:1.*

Josh: *No...* ***I've*** *dropped out.*

Karen: *He hasn't dropped out formally. He's just stopped going.*

David: ***Why?***

2. In this section, we use 'transcript' rather than 'subtitles' to align with the terminology used in the pedagogical literature (see e.g. Washburn 2001).

Josh: ***I'm*** *not happy, Dad.*

David: ***Happy? Who cares about happy? You*** *go to university to get a degree.*

Josh: ***I'm*** *no good at Spanish.* ***You*** *saw that for* ***yourself****. And* ***I*** *hate business studies.* ***I*** *only did it to please* ***you****.*

David: *So dropping* ***out's*** *gonna make* ***me*** *proud?*

Firstly, in groups of three (matching the number of characters in the extract), students will be shown the short (47 second) clip corresponding to this extract without sound.[3] While students may not guess the exact topic of the conversation, they can be expected to observe the domestic setting, may discuss the possible relationships between the characters, and comment on the apparent tension. Groups will need to be given enough time to discuss these observations. In step 2, roles can be allocated by the teacher or within the groups themselves. To help the groups prepare their dialogues, they can be given a handout with a template listing the sequence of the characters' turns with the utterances blanked out. When watching the clip again in step 3, now for the first time with sound, students can use the handout to identify any differences between their version of the dialogue and the original. The written format is emphasised in step 4, when students are given a handout with the full transcript. At this point, students can be asked to underline the turns that indicate differences in expectations. These steps should prepare students well to voice the dialogue without the aid of the transcript in the final step. Depending on the length of the lesson, the activity can be concluded with a discussion about the role of parents in decisions about university and career choices in the students' own cultural contexts.

In general, the clip chosen for the activity will need to be quite short so that sufficient class time is available to cover the steps outlined above for learners to meaningfully engage with and critique the clip (see Section 2.2). Step 3, in particular, allows students to "critically compare their own performances" (Anderson & Corbett 2009: 177) to the *Cold Feet* scene. The choice of scene can be guided by a feature of natural conversation searched via the *Cold Feet* corpus. Moreover, students could themselves directly engage with the corpus data (for examples of lesson plans involving hands-on corpus activities, see e.g. Viana 2023).

3. Readers at UK educational institutions may access this clip via the Box of Broadcasts on demand television service for education (ITV London 2019). Alternatively, IMDb hosts several publicly available clips (*Cold Feet* 1997).

6. Conclusions

The present study aimed to examine to what extent extracts from subtitles of the British comedy drama *Cold Feet* could serve as a classroom model for teaching English language speaking by exposing students to useful elements of natural conversation. In response to the first research question, this study largely supports the claim that the language in *Cold Feet* closely resembles that of natural conversation. The results of the additive MDA showed that, in the three dimensions that highlight the differences between spoken and written discourse, the language in *Cold Feet* falls towards the spoken end of the continuum. The analysis of individual linguistic features using 3-grams showed a high rate of features that typically occur in spoken texts. However, the MDA showed that the language in *Cold Feet* does not map exactly onto natural conversation and in two dimensions it aligned most closely with "general fiction". These results are perhaps not surprising considering the fact that *Cold Feet* is scripted. Additional salient reasons causing the differences between *Cold Feet* and natural conversation are that the main communicative purpose of the language in *Cold Feet* is to tell a narrative, and to entertain, and that it is not produced in real-time. These situational factors account for the lack of some features that we would expect to find in naturally occurring conversation such as overlaps and interruptions.

This study only examined ten episodes of one television programme, so it is not possible to draw conclusions about all telecinematic language. However, from the results found here, it can be said that if extracts from these episodes of *Cold Feet* were used as English language classroom models, the students would be exposed to many salient features of natural conversation in British English. Future research could examine whether Cooper's (2023) methodology for using MAT combined with cluster analysis could be adapted to help identify extracts based on specific language features. Furthermore, we would welcome pedagogically oriented studies focusing on telecinematic language beyond the US (which have tended to be the focus of prior research) and the UK (as in the present study) to aid the development of pop culture materials in the spirit of the Global English Language Teaching paradigm (see e.g. Bagni & Sumra 2022).

Funding

Open Access publication of this article was funded through a Transformative Agreement with the University of Stirling.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful feedback, the editorial team, MAT developer Andrea Nini, and the Sketch Engine Team.

References

Al-Surmi, M. (2012). Authenticity and TV shows: A multidimensional analysis perspective. *TESOL Quarterly*, *46*(4), 671–694.

Al-Surmi, M. (2022). TV Shows, authenticity, and language learning: A corpus-based case study. *Register Studies*, *4*(1), 30–54.

Anderson, W., & Corbett, J. (2009). *Exploring English with online corpora: An introduction*. Basingstoke: Palgrave Macmillan.

Bagni, M., & Sumra, Z. (2022). Beyond monolithic English: Pop culture materials informing Global English Language Teaching. *International Journal of Applied Linguistics*, *173*(2), 286–310.

Barber, N. (1998, November 22). Comedy drama: What's less believable than Dr Who? *Independent*. Retrieved from https://www.independent.co.uk/arts-entertainment/comedy-drama-what-s-less-believable-than-dr-who-1186551.html

Basturkmen, H. (2001). Descriptions of spoken language for higher level learners: The example of questioning. *ELT Journal*, *55*(1), 4–13.

Bednarek, M. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge: Cambridge University Press.

Berber Sardinha, T. (2014). 25 years later: Comparing internet and pre-internet registers. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 81–105). Amsterdam/Philadelphia: John Benjamins.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1989). A typology of English texts. *Linguistics*, *27*(1), 3–43.

Biber, D. (2023). Register and discourse analysis. In M. Handford & J. P. Gee (Eds.), *The Routledge handbook of discourse analysis* (pp. 309–327). London: Routledge.

Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at…*: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Cold Feet [Comedy, Drama, Romance]. (1997). Granada Television. Retrieved from https://www.imdb.com/title/tt0168596/

Cooper, C. R. (2023). The identification of YouTube videos that feature the linguistic features of English informal speech. *Applied Corpus Linguistics*, *3*(3), 100068.

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 197–216). Amsterdam/Philadelphia: John Benjamins.

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume*. Council of Europe Publishing. https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-descriptors

Csomay, E., & Petrović, M. (2012). "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System*, *40*(2), 305–315.

Forchini, P. (2012). *Movie language revisited: Evidence from multi-dimensional analysis and corpora*. Bern: Peter Lang.

Forchini, P. (2021). Movie discourse: Marvel and DC Studios compared. In E. Friginal & J. A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 183–201). London: Routledge.

Gregory, M., & Carroll, S. (1978). *Language and situation: Language varieties and their social contexts*. London: Routledge & Kegan Paul.

ITV London. (2019, 20 January). Clip from *Cold Feet*, 23:20, 55 mins. 00:21:54-00:22:41. Retrieved from https://learningonscreen.ac.uk/ondemand/index.php/clip/255169?bcast=128402327

Jones, C. (2017). Soap operas as models of authentic conversations: Implications for materials design. In A. Maley & B. Tomlinson (Eds.), *Authenticity in materials development for language learning* (pp. 158–175). Newcastle: Cambridge Scholars Publishing.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36.

Kozinski, R.A. (2011). Quantifying the emotional tone of James Bond films. In R. Piazza, M. Bednarek, & F. Rossi (Eds.), *Telecinematic discourse: Approaches to the language of films and television series* (pp. 125–140). Amsterdam/Philadelphia: John Benjamins.

Lazzeretti, C. (2019). 'I've been nowhere and done nothing'. The characterization of Daisy Mason in the British drama *Downton Abbey*. *Series — International Journal of TV Serial Narratives*, *5*(1), 33–44.

Le Foll, E. (2021). Register variation in school EFL textbooks. *Register Studies*, *3*(2), 207–246.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, *22*(3), 319–344.

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, *55*(2), 163–192.

Nini, A. (2019). The multi-dimensional analysis tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67–94). London: Bloomsbury Academic.

Nini, A. (2020). *Multidimensional Analysis Tagger (v. 1.3) — Manual*. Retrieved from https://sites.google.com/site/multidimensionaltagger/

Nini, A. (2022). *Multidimensional Analysis Tagger (Version 1.3.3 for Windows)*. Retrieved from https://sites.google.com/site/multidimensionaltagger

Piazza, R., Bednarek, M., & Rossi, F. (2011). Introduction: Analysing telecinematic discourse. In R. Piazza, M. Bednarek, & F. Rossi (Eds.), *Telecinematic discourse: Approaches to the language of films and television series* (pp. 1–17). Amsterdam/Philadelphia: John Benjamins.

Quaglio, P. (2009). *Television dialogue: The sitcom* Friends *vs. natural conversation*. Amsterdam/Philadelphia: John Benjamins.

Rotten Tomatoes. (2023). *Cold Feet*. Retrieved March 31, 2023, from https://www.rottentomatoes.com/tv/cold_feet

Subslikescript. (2024). *Cold Feet*. Retrieved August 12, 2024, from https://subslikescript.com/series/Cold_Feet-168596

Thornbury, S., & Slade, D. (2006). *Conversation: From description to pedagogy*. Cambridge: Cambridge University Press.

Toolan, M. (2011). "I don't know what they're saying half the time, but I'm hooked on the series": Incomprehensible dialogue and integrated multimodal characterisation in *The Wire*. In R. Piazza, M. Bednarek, & F. Rossi (Eds.), *Telecinematic discourse: Approaches to the language of films and television series* (pp. 161–183). Amsterdam/Philadelphia: John Benjamins.

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 173–180. Retrieved from https://aclanthology.org/N03-1033.

Veirano Pinto, M. (2014). Dimensions of variation in North American movies. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (pp. 109–147). Amsterdam/Philadelphia: John Benjamins.

Veirano Pinto, M. (2018). Variation in movies and television programs: The impact of corpus sampling. In V. Werner (Ed.), *The language of pop culture* (pp. 139–161). London: Routledge.

Viana, V. (Ed.). (2023). *Teaching English with corpora: A resource book*. London: Routledge.

Vu, D. V., & Peters, E. (2022). The role of formulaic sequences in L2 speaking. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 285–298). London: Routledge.

Washburn, G. N. (2001). Using situation comedies for pragmatic language teaching and learning. *TESOL Journal*, *10*(4), 21–26.

Werner, V. (2021a). A diachronic perspective on telecinematic language. *International Journal of Corpus Linguistics*, *26*(1), 38–70.

Werner, V. (2021b). Catchy and conversational? A register analysis of pop lyrics. *Corpora*, *16*(2), 237–270.

Werner, V., & Tegge, F. (2020). Introduction: Learning languages through pop culture/learning about pop culture through language education. In V. Werner & F. Tegge (Eds.), *Pop culture in language education*. London: Routledge.

Wilson, J. J. (2008). *How to teach listening* (11th ed.). Harlow: Pearson.

Address for correspondence

Viola Wiegand
Faculty of Social Sciences
RG Bomont
University of Stirling
Stirling FK9 4LA
Scotland, UK
viola.wiegand@stir.ac.uk
https://orcid.org/0000-0001-6746-6620

Co-author information

Jill Cannon-Jones
University of Stirling
jillcannonjones@gmail.com
https://orcid.org/0009-0007-0458-206X

Publication history

Date received: 14 August 2024
Date accepted: 24 April 2025
Published online: 28 July 2025