"Get us the hell out of here"

Key words and trigrams in fictional television series

Monika Bednarek
University of Sydney, Australia

Based on the analysis of key words and trigrams, this paper explores characteristics of contemporary American English television dialogue. Using a corpus comprising dialogue from seven fictional series (five different genres) and the spoken part of the American National Corpus, key words and trigrams are compared to previous corpus linguistic studies of such dialogue (Mittmann 2006, Quaglio 2009) and further explored on the basis of concordances, with special emphasis on over-represented key words/trigrams and their potential to indicate informality and emotionality. The results suggest that the expression of emotion is a key defining feature of the language of television, cutting across individual series and different televisual genres.

Keywords: television dialogue, key words, n-grams, scripted language, television series

1. Introduction

This paper explores American English television dialogue from a corpus linguistic perspective. While television dialogue has, to date, only rarely been the object of corpus linguistic inquiry, it has recently emerged as an important new area of research for pragmatics, sociolinguistics, conversation analysis, translation studies and stylistics. Such linguistic analysis has shown that television dialogue offers manifold areas of interest, ranging from the linguistic construal of characterisation and character relations (Bednarek 2011c, Bubel 2006, Mandala 2007, Richardson 2010) to issues of translation (Díaz-Cintas 2009, Mittmann 2006), narrative, humour, ideology and multimodality (e.g. contributions to Piazza et al. 2011; Stokoe 2008). A principle focus of this research has been on the analysis of characterisation through language. For example, Bubel & Spitz (2006) show how verbal humour is an important means in *Ally McBeal* (Fox, 1997–2002) of characterizing

International Journal of Corpus Linguistics 17:1 (2012), **35–63**. DOI 10.1075/ijcl.17.1.02bed
ISSN 1384–6655 / E-ISSN 1569–9811 © John Benjamins Publishing Company

televisual characters, and Mandala (2007) demonstrates how marked *-y* suffix adjectives in *Buffy the Vampire Slayer* (Warner Brothers, 1997–2003) work in the construal of televisual friendship groups. Bubel (2006) explores how alignment patterns, terms of address and questions contribute to the linguistic construal of character relations in *Sex and the City* (HBO, 1998–2004), and Richardson (2010) draws on impoliteness theory to investigate the character House in the medical drama of the same name.

Corpus linguistic studies of television dialogue have also tended to focus on individual series. Rey (2001) explores dialogue from various *Star Trek* series (NBC/Paramount, 1966–1969, 1987–1994, 1993–1999) with a focus on gender roles; Mittmann (2006) analyses data from *Golden Girls* (NBC, 1985–1992), *Dawson's Creek* (Warner Brothers, 1998–2003) and *Friends* (NBC, 1994–2004) in the context of "translationese"; Baker (2005) explores the construction of identities in *Will & Grace* (NBC, 1998–2006); Quaglio (2008, 2009) compares the language of *Friends* (NBC, 1994–2004) to American conversation using multidimensional analysis and Bednarek (2011a, b) investigates various aspects of dialogue in *Gilmore Girls* (Warner Brothers, 2000–2007).

However, neither studies in or outside corpus linguistics have so far systematically explored the question to what extent television dialogue is a language variety in its own right and what features would characterise such a variety (but see Bednarek 2010 on limited TV data). This paper offers a first investigation into this issue using a corpus comprising dialogue from 32 episodes in seven television series. Key word/n-gram analysis (Scott & Tribble 2006) is employed to compare this scripted dialogue with unscripted spoken American English. There are two main reasons why this methodology was chosen to explore television dialogue: First, it has been shown to discriminate text types, and second, it has already been successfully applied to scripted data, including television dialogue. In the rest of this paper, I first introduce this methodology and the corpora to which it has been applied (Section 2) before discussing the results (Section 3). I conclude by drawing together the analyses and pointing to areas delineated for future research (Section 4).

2. Description of methodology and corpora

2.1 Key word/n-gram analysis

'Keyword' analysis is a method that uses corpus linguistic software such as Wordsmith (Scott 2008) for identifying linguistic features that are unusual in terms of frequency when a "node" text or corpus is compared to a reference corpus (see Scott &

Tribble 2006:59). The linguistic features that are unusually (in)frequent in the node corpus are identified by the software as 'key', pointing the researcher to features that characterise the language of the corpus. These linguistic features can be individual word forms (e.g. *is, be, was*) — i.e. key *words* — or they can be longer multi-word structures (e.g. *you know*, *a lot of*, *in the middle of*) — i.e. key *clusters* or *n-grams*. N-grams can be defined as "multi-word strings of two or more uninterrupted word-forms" (Stubbs & Barth 2003:62), that is, longer syntagmatic structures rather than single words or word forms. Thus n-grams include bigrams (e.g. *you know*), trigrams (e.g. *a lot of*), etc. Following corpus linguists such as Biber et al. (1999) these n-grams can be identified on a purely automatic, quantitative basis, i.e. without an a-priori consideration of meaning. This means that identified n-grams can be "fragmentary strings", "meaningful strings" and semantic/pragmatic expressions (O'Keeffe et al. 2007:61), and it is not claimed that they are linguistic units with grammatical, semantic or pragmatic status (Stubbs & Barth 2003:69; see further Biber 2009 on the typical structure and functions of n-grams). However, even fragments are "important clues as to how interaction unfolds" (O'Keeffe et al. 2007:70).

N-grams have been studied under various names (dyads/tryads, clusters, phrases, lexical bundles, n-grams, chains) in corpus linguistic analyses (e.g. Biber et al. 1999; Mittmann 2004, 2006; O'Keeffe et al. 2007; Stubbs & Barth 2003). These analyses have revealed, *inter alia*, that n-grams are register/genre/text-type sensitive (e.g. Biber 2009, Culpeper & Kytö 2010, O'Keeffe et al. 2007:61–62, Stubbs & Barth 2003). For instance, Stubbs & Barth's (2003) analysis of three corpora representing fiction, belles lettres and academic writing shows that these text types are distinguished both by the n-grams that occur as well as by their repetitiveness. Thus, a trigram such as *I don't know* (note that the software counts *don't* and similar contractions as one word) is highly frequent in fiction but not in academic writing. Stubbs & Barth (2003:69) conclude that n-grams can "discriminate between text-types". Another reason for applying this methodology concerns the fact that both key *word* and key *n-gram* analyses have been successfully applied in exploring other types of scripted language such as drama texts (e.g. Culpeper 2009), and, crucially, the methodology has also been employed in the study of television dialogue (e.g. Baker 2005, Mittmann 2006).

In this paper I will discuss the results for analysing both key *words* as well as key *trigrams* (combinations of three words) in television dialogue. Key words are useful because they are both highly frequent and well dispersed across the corpus. Trigrams are reasonably frequent in my corpus and allow slightly better hypotheses as to their function than bigrams do (which are even more frequent). For instance, bigrams such as *what the*, *need to*, *out of*, *are you* are less informative as to their potential function than are trigrams such as *what the hell*, *need to talk*, *out of here*, *what are you*. Thus, trigrams have the advantage of being frequent enough in my

corpus to warrant analysis and discussion as well as indicating functional meanings more unambiguously (though not totally unambiguously, as will be seen).

To sum up, a key word/n-gram analysis compares the frequency of words and n-grams in a given corpus with that of a reference corpus and results in a list of key words and n-grams that are more or less frequent in this corpus when compared to the reference corpus. For this paper a corpus of scripted dialogue from American English fictional television is compared to a reference corpus comprising unscripted spoken American English.[1] The following section introduces both corpora.

2.2 Corpora

2.2.1 *The television corpus ("TV corpus")*

The node corpus used in my study of key words and trigrams consists of dialogue from seven different American English fictional television series. To compile this corpus I made use of online fan transcripts (not the original scripts). Such transcripts have also been used in previous corpus linguistic studies such as those by Baker (2005; personal communication) and Quaglio (2008, 2009). Quaglio (2008:191–192) notes that the fan transcripts for *Friends* that he uses are "fairly accurate and very detailed, including several features that scripts are not likely to present: hesitators, pauses, repeats, and contractions".

The corpus compiled for this study (henceforth: the TV corpus) does not contain any of the series investigated in previous corpus linguistic studies of television dialogue (i.e. *Star Trek*, *Will & Grace*, *Dawson's Creek*, *Golden Girls*, *Friends* — cf. Baker 2005; Mittmann 2006; Quaglio 2008, 2009; Rey 2001), as the aim was to analyse more contemporary series. The corpus was also designed to be more varied with respect to television genres, aiming specifically and intentionally to include a range of diverse genres. Thus, it contains 32 episodes from seven series representing different genres such as crime, mystery, medical drama, comedy and drama, with 24,000 to 26,000 words for each genre (130,000 words in total). This variety is necessary to ensure that the corpus is not just representative of specific television genres such as comedy or drama or specific television series such as *Star Trek* or *Friends*. Table 1 shows the episodes and series included in the corpus.[2]

2.2.2 *The spoken American English corpus ("ANCS")*

The reference corpus to which this corpus was compared is the spoken part of the second release of the *American National Corpus* (ANCS). This corpus consists of a variety of unscripted spoken American English (as described on the ANC website at http://www.americannationalcorpus.org): unscripted telephone conversations (3,019,477 words), narratives (198,295 words, from the Charlotte Narrative and Conversation Collection) and academic spoken English (593,288 words, from the

Table 1. The television corpus

Genre	Series	No of tokens according to WordSmith
Crime	*NCIS* (episodes 1.01, 2.06, 2.09, 2.10, 2.13)	26,601
Mystery	*Supernatural* (episodes 1.10, 1.14, 1.18, 1.19) *Lost* (episodes 1.11, 1.15, 1.17)	25,553
Medical	*House* (episodes 1.02, 1.08, 1.12, 1.14, 1.18)	26,697
Comedy	*How I Met Your Mother* (episodes 1.04, 1.06, 1.08, 1.10, 1.12) *My Name is Earl* (episodes 1.03, 1.07, 1.12, 1.17, 1.21)	25,406
Drama	*Desperate Housewives* (episodes 1.02, 1.07, 1.11, 1.15, 1.19)	24,032
Total	Seven series, 5 genres, 32 episodes	128,289

Michigan Corpus of Academic Spoken English). In total, the WordSmith software (Scott 2008) used in this study counts the number of running tokens for the spoken part of the ANC as 3,764,279. While the ANCS is not a perfect reference corpus because it includes what I presume is elicited data (the Charlotte narratives), and is mainly made up of telephone conversations, it represents not only *spoken* but also *unscripted* American English. As the elicited data comprise less than 6% of the overall corpus, they are unlikely to influence the findings much.

3. Key words and trigrams in seven fictional television series

Even with an extremely low *p*-setting (see Note 1) Wordsmith identified hundreds of key words and key trigrams. It was beyond the scope of the research project to analyse all of these. Thus, I only considered the top 30 positive and the top 30 negative key words, excluding names (e.g. *Booth, NCIS*) and symbols (#). (Unsurprisingly, the top 24 positive key words are just names of televisual characters.) For key trigrams I considered all 39 "negative" trigrams and among the top 39 "positive" trigrams I focused on those that occur in the TV corpus with a frequency of at least 0.01 (frequency as a percent of the total as calculated by Wordsmith) to ensure some dispersion. Again, I excluded names (e.g. *Air Force One*) and symbols. The resulting positive and negative key words and key trigrams are listed in Table 2.

Table 2. Key words and trigrams in the TV corpus

	"Positive" (unusually frequent)	**"Negative" (unusually infrequent)**
Key words	*gonna, gotta, hell, hey, mom, dad, guy, boss, man, ok, him, his, her, he, thanks, sorry, please, Mr, Mrs, Dr, agent, detective, why, me, I'm, your, you're, this, no*	*so, really, but, because, kind, uh, um, huh, know, think, well, things, then, had, and, it, they, they're, them, I, that, that's, yeah, of, or, like, people, have, in, lot*
Key trigrams	*no no no, what the hell, the hell is, oh my god, need to talk, out of here, I love you, I can't believe, I want you, you want me, want me to, what are you, why don't you, who are you, why are you, are you doing, are you okay, I told you*	*I I I, I I don't, yeah yeah yeah, kind of a, it's kind of, uh I don't, uh you know, I mean I, I mean you, you know I, that you know, and you know, you know you, you know the, you know it's, don't know I, yeah you know, I don't know, you know we, don't know if, you know it, but you know, uh you know, I think that, and I think, yeah I think, I think that, I think it, I think the, I think it's, I don't think, yeah well I, something like that, and it was, and then you, a lot of, a little bit, some of the, one of the*

As mentioned, the TV corpus is comprised of the language of seven different television series (*NCIS*, *Supernatural*, *Lost*, *House*, *How I Met Your Mother*, *My Name is Earl*, *Desperate Housewives*). In order to investigate whether or not television dialogue is a language variety in its own right, it is useful, as a first step, to compare the findings in Table 2 for the seven series in my corpus with previous corpus linguistic findings for three *different* television series (Mittmann 2006, Quaglio 2008, 2009). Mittmann (2006) used a corpus comprising 14 episodes from three television series: seven episodes from *Friends* (17,099 words), six episodes from *Dawson's Creek* (29,416 words) and one episode from *Golden Girls* (3,109 words), thus only including the two genres of sitcom (*Friends*, *Golden Girls*) and drama (*Dawson's Creek*). While her study focuses on translationese, she does apply key word/n-gram methodology to compare this corpus with a corpus of unscripted conversation and briefly talks about over-and under-represented word forms and n-grams. Quaglio (2008, 2009) uses a much bigger corpus (600,000 words) but only investigates the language of the sitcom *Friends*, employing Biber's (1988) multidimensional analysis to compare it to unscripted conversation. Note that both Mittmann and Quaglio explore *Friends*, which means that we may know a lot about TV dialogue in this particular series, but not so much about other, more contemporary television fiction.

The argument here is that if there are commonalities between my findings (for seven television series representing five genres) and Mittmann's and Quaglio's findings (for three different series representing two genres), they may point to linguistic features that are characteristic for television dialogue as a whole, rather than only specific television series (*Friends*) or genres (sitcom, drama). This is one of the advantages of this comparison — another is that it will pinpoint linguistic features worth exploring further.

3.1 Specific key words/trigrams

If we compare the results reported in Mittmann (2006) and Quaglio (2009) with the results in Table 2 we can see that some linguistic features appear to be shared by different television series. Table 3 shows that there are indeed some word forms and trigrams that are "key" in my corpus which have also been identified by either Quaglio (2009) — identified as (Q) in the table — as being more frequent in *Friends* than in conversation or by Mittmann (2006) — identified as (M) in the table — as being more frequent in her corpus than in unscripted conversation.

Table 3. Key words and trigrams compared to other studies of TV dialogue

Key words	Key trigrams
gonna (Q), *gotta* (Q), *hell* (M), *hey* (M & Q), *man* (Q: as vocative[3]), *thanks* (M), *sorry* (M), *please* (M), *me* (Q: first-person pronouns), *I'm* (Q: first-person pronouns), *your* (Q: second-person pronouns), *you're* (Q: second-person pronouns)	*no no no* (Q: repeats, e.g. *I–I-I*), *what the hell* (M), *the hell is* (M), *oh my god* (M), *I can't believe* (Q)

Thus, Mittmann notes that her television corpus has more "greeting expressions" including *hey*, more "interactive markers" including *please* and *thanks*, and more "apologizing expressions" including *I'm sorry*. Her corpus also has more of the expletives *god* and *hell* compared to conversation, with *hell* occurring mostly as part of the phrase *WH- + the hell* (predominantly *what/who/how the hell*). Quaglio (2009:158–161) specifically lists first- and second-person pronouns, the vocative *man* (and others), the trigram (or lexical bundle, as he calls it) *I can't believe*, the greeting *hey* (and others), repeats (e.g. *I–I-I*) and semi-modals (but no specific numbers for *gotta/gonna*) as being more frequent in *Friends* than in conversation. The expressions in Table 3, then, are potential candidates for a "common" lexical and grammatical vocabulary for television dialogue. It must be pointed out, though, that neither Mittmann nor Quaglio use the ANCS as reference corpus — rather, they use a conversation subcorpus of the Longman Grammar Corpus (to

which I had no access). Thus, some differences might be the result of the specific nature of the respective reference corpus, even though Scott & Tribble (2006:64) suggest that "while the choice of reference corpus is important, above a certain size, the procedure throws up a robust core of KWs [key words] whichever the reference corpus used".

We also need to go further with this comparison. Considering Table 3, it is not enough to say that the listed words and trigrams are key in the television corpus, we also need to take into account dispersion. In other words, do these key words and trigrams occur in *all* episodes and in *all* series? Table 4 provides a summary of both raw frequency and dispersion across episodes and series.

Table 4. Frequency and dispersion

Word/trigram	Raw frequency	Dispersion (episodes)	Dispersion (series)
me	976	32/32	7/7
I'm	972	32/32	7/7
your	789	32/32	7/7
you're	507	32/32	7/7
hey	301	32/32	7/7
sorry	195	32/32	7/7
gonna	408	31/32	7/7
man	170	30/32	7/7
thanks	92	29/32	7/7
please	102	28/32	7/7
gotta	93	27/32	7/7
hell	79	26/32	7/7
the hell is	14	12/32	7/7
no no no/no, no, no	39	15/32	6/7
what the hell	22	15/32	6/7
I can't believe	19	15/32	6/7
oh my god/oh, my god	34	7/32	6/7

As Table 4 shows, the "core" features occurring across at least 80% of all episodes and in all seven series are made up of key word forms involving first- and second-person pronouns (*me, I'm, your, you're*), as well as the specific key words *hey*, *sorry*, *gonna*, *gotta*, *man*, *thanks*, *please* and *hell*. These core linguistic features, then, are both shared with other TV series and well-dispersed. They can be linked to informality, the expression of emotion, involvement and specific communicative situations as follows:

- *Hey*, *man* (as vocative), and semi-modals have been associated with informality, i.e. colloquial language (e.g. Quaglio 2009, following Biber 1988);
- The expletive *hell* can be associated with the expression of emotion (e.g. Bednarek 2010);
- First- and second-person pronouns have been linked to discourse immediacy. Discourse immediacy is defined as a "focus on immediate concerns, facts, and evaluative utterances" (Quaglio 2009:137), including a focus on the "involvement of the interaction" (Quaglio 2009:133).
- Routine formulae like *sorry*, *thanks* and *please* are associated with particular communicative situations (Mittmann 2006).

I will investigate informality and the expression of emotion in much more detail below, but briefly, informal, or colloquial language might be used to hide the scripted nature of television dialogue. Quaglio (2009:120) argues that it makes televisual language "credible and authentic". In other words, it may contribute to the "willing suspension of disbelief" on the part of audiences and allow them to engage with what they perceive as "realistic" characters, characters that are realistic because they use the same kind of colloquial language as viewers do. At the same time, it reflects character relationships (Quaglio 2009:120) in that friends, family members, etc. use informal language with each other — more on this point later (Section 3.2).

The expression of emotion (emotionality), on the other hand, can be associated with the presence of "dramatic" or "emotional" moments in television series that seem to keep audiences interested. This issue will be discussed in more depth further below (Section 3.3). As far as discourse immediacy or a focus on interactive character involvement is concerned, it is obvious that much television dialogue focuses on the speaker-hearer relationship, reflecting the fact that many series are centred on relations between people. A series such as *House* for instance, is characterized by the witty dialogue between characters, arguably aimed at amusing the audience, as are other fictional television series such as *Gilmore Girls* (Bednarek 2010) and *Sex and the City* (Bubel 2006:42). Consider the following "House-isms" from the official House website:

(1) CUDDY: Dr. House! Need **you** here.
HOUSE: No thanks. Lotta sick people. **I** might catch something.

(2) CUDDY: **You**. In the lobby. Now.
HOUSE: **I** hurt my leg. **I** have a note.

As these "House-isms" show, the witty exchanges between two characters, as commands are issued and insults are thrown (see Richardson 2010 for an investigation

of impoliteness in *House*), necessarily involve the use of first and second-person pronouns referring to speaker and hearer respectively.

Finally, particular communicative situations that necessitate the use of specific routine formulae may be more frequent in TV dialogue because television series are structured around frequent changes of different communicative situations. At the same time, at least two of the specific formulae identified here (*sorry*, *thanks*) can be linked to the conventionalised expression of the emotions of gratitude and regret and may thus be associated with emotionality.

Following this general exploration of a potential "core" lexical and grammatical vocabulary for television dialogue, a more detailed analysis is now offered of two of the functional dimensions described above — informality and emotionality. These two were chosen for further investigation because most of the identified "positive" key words and trigrams can be discussed in such terms.

3.2 How "informal" is television dialogue?

In his analysis of *Friends*, Quaglio (2009) argues that the language of *Friends* is more informal or colloquial than natural conversation, as evidenced by higher frequencies for expletives, slang terms, vocatives, innovations, semi-modals, repeats and the greetings/leave-takings *hi, hey, bye* and *bye-bye*. So is television dialogue in general (rather than just dialogue in the sitcom *Friends*) more informal than unscripted conversation? To investigate this further, I applied Quaglio's categorisation to the results for key words and trigrams (as listed in Table 2 above). As I will show, the results suggest a more nuanced answer, rather than a simple "yes" or "no".

On the one hand, the television corpus does appear more "informal" than the spoken corpus: Two semi modals (*gonna, gotta*), the expletive *hell* — including the trigrams *what the hell*, *the hell is* — as well as the informal *hey*, and the informal words *mom, dad, guy,* and *boss* are over-represented in the television corpus.[4] On the other hand, three repeats (*I I I, I I don't, yeah yeah yeah*) are under-represented, while one repeat is over-represented (*no no no*). Repeats thus deserve a closer look. Quaglio (2009) notes that together with hesitations and false starts, "repeats are one of the features resulting from the pressures of online production" (Quaglio 2009:118), and certainly at least two of the repeats that are under-represented in the television corpus (*I I I, I I don't*) seem to be the result of the online cognitive effort of unscripted conversation, which takes place in real time. In contrast, as Figures 1 and 2 on pages 45–46 indicate, the repeat *no no no* that is **over**-represented in the television corpus is tied more to the expression of emotionality (emphasis, urgency) or contradiction.

In both of these figures, the squares around particular parts of the concordances point to features that allow us to tie the repeat *no no no* both to emphatic

contradiction (compare the frequent negated statements such as *it's not gonna happen*) and to the expression of emotionality, e.g. interjections such as *oh*, exclamation marks, expletives like *oh crap,* apologies, imperatives such as *listen to me, just take a breath, don't please*.

So one area that needs further research concerns the question whether non-fluency features such as repeats should be treated as indicators of emotionality rather than informality. In contrast to unscripted language, television dialogue is scripted beforehand and any non-fluency features would either be specifically included by scriptwriters as a clue to a character's state-of-mind or be the outcome of an actor's performance indicating this state-of-mind. This partially holds for other drama texts: "if features normally associated with normal non-fluency do occur, they are perceived by readers and audience as having a *meaningful* function precisely because we know that the dramatist must have included them *on purpose*" (Short 1996:177 cited in Culpeper & Kytö 2010:145–146). However, if we are looking at television transcripts, these are the outcome of a *performance* by actors so that multiple authorship for such features is possible: the television scriptwriters or the actors/directors. Quaglio (2009:119) notes that "actors are sometimes directed to express certain types of behaviour or attitude that transcend the scripted dialogues (e.g. reluctantly, excitedly); they thus need to make certain linguistic choices to convey these behaviours or attitudes, and repeats can be instrumental for the achievement of this dramatic effect". An illustration of this can be seen in example 3, where a character (Ally) performs a joke in front of a big audience:

(3) ALLY: MEN can handle it [telling dirty jokes],
women can't we're we're uh
we're not tough enough.
we're we're too we're too (2.0) fragile.
(*Ally McBeal*, transcription from Bubel & Spitz 2006)

As Culpeper (2001:191) notes, repetition can mark emotional anxiety and this is what seems to be the case here: The false starts and repeats appear to signal Ally's nervousness at standing up in front of an audience and confronting it with

1 Yes, I want Tahitian vanilla. You wrote it down wrong? No no no no, listen to me. If I go to my wedding and the
2 : I don't know, but we have plenty of chocolate here. Ted: No no no you don't understand. I need that Kit Kat. She
3 ideas. I'm sorry-I'm sorry, I didn't mean that. TOM: No, no no no no. We both know that your career was going so
4 : Hello Randy: I want to start out by saying I'm sorry. Earl: no no no no no Randy did you tell her where we were.
5 a baby body? Oh crap oh crap oh crap. Baby killer. Earl: no no no no no, just take a breath. Everythings cool I just
6 to be in solitary then. Is that what you saying? Earl: no no no that's not what I'm saying. Earl: you don't
7 that Earl? What is that? Jesse: Jeff call the cops. Earl: no no no no don't please. Don't call the cops. I-i-I'm trying to
8 the cord. Earl: Randy! Jesse: We're gonna die. Earl: no no no no no I'm sorry - look you're not gonna die. I'm sorry.
9 called Xena: I love you too Earl. you'll be my first. Earl: no no no no I was talking about karma. Xena: oh karma. Earl:

Figure 1. Example concordances for *no no no*

1 all down on a cocktail napkin. DEAN: Not me. SAM: Oh, no, no, no, no. Pickups are your thing, Dean. DEAN: It
2 : Uh, a VW? GIBBS: It's a car, McGee! MCGEE: Ah, no, no, no. I know, I just didn't know that you owned a VW.
3 You mean like physically abusing him? McGee: No. No, no, no, no. No, Gibbs never lays a finger on a suspect.
4 go back to how we were before she came home. David: No, no, no, Greta, it's not gonna happen, okay? We were
5 MICHAEL: Say it again! JACK: Michael! MICHAEL: No, no, no, say it again, man! JACK: Michael! Michael!
6 : We love tiramisu. Am I wrong in saying that. ? Robin: No, no, no, I mean it just sounds a little bit weird, doesn't
7 : TRY! Ted: It's… ineffable. Natalie: I'm not "F-able"? Ted: no, no, no, no. Ineffable, ineffable means it can't be
8 Natalie: Oh, so dating you is like winning the lottery? Ted: No, no, no, I didn't mean that. Natalie: Okay, So what's
9 I thought you were going after an in-house position. Tom: No, no, no, I'm setting up new offices from Seattle to L.A. I,
10 fact, I just think of you as a really good buddy. Susan: No. No, no, no, You obviously think you have some insight
11 breaking up with me. In my own bedroom. GABRIELLE: No, no, no, I'm not dumping you. Lately, you've been the
12 LYNETTE: Well, bring 'em, we're leaving. LADY: Oh, no, no, no. Stay put. LYNETTE: Oh, I don't think so. Oh!
13 : Don't be that way. I got you a gift. GABRIELLE: Nope. No, no, no, no, you're not gonna buy your way outta this
14 us all. House: Thank you. House: So, sing for me. Oh, no, no, no, no… come on, look. When you had your surgery,

Figure 2. Example concordances for *no, no, no*

gender stereotypes. The confrontational nature of Ally's mini-lecture here is noted by Bubel & Spitz (2006:91).

Indeed, I would hypothesise that further research will show that those non-fluency features that do occur in television dialogue can often be tied to the expression of emotion (e.g. nervousness) and that their key function is to convey a character's state of mind rather than to signal informality. This hypothesis seems likely given that in other scripted texts — Early Modern English Drama comedy — Culpeper & Kytö (2010) found that only a minority of lexical repeats can be classified as "performance 'errors'" (Culpeper & Kytö 2010:147), and many more are tied to the expression of agreement, disagreement, intensification and expressivity.

Another issue relevant to the level of informality in television dialogue concerns the use of titles, which arguably signal formality rather than informality. In fact, as shown in Table 2 various titles feature among the over-represented key words in the TV corpus (*Mr, Mrs, Dr*, certain uses of *agent, detective*). Such titles are frequent on the one hand, because specific television series and genres feature agents and detectives (cf. Figures 3 and 4). As concordance lines 3 and 4 in Figure 4 show, the prevalence of *detective* as a key word may be the result of the corpus including the names of characters as speakers (compare Note 2), with certain minor characters simply being labelled by their title. If we just consider the *interactive* use of titles (used to talk about oneself or address another, e.g. *I'm Dr/Mr/Mrs/Agent/Detective X — Dr/Mr/Mrs/Agent/Detective X, can you…*), we can see that *agent* and *detective* have a very low dispersion across the corpus (Table 5). In fact, there is a drop in dispersion across all titles, with dispersion across **episodes** ranging from 3% (*detective*, which occurs in only one episode) to 47% (*Mr*, which occurs in 15 episodes). While *Mr* shows a dispersion across 5 of the 7 **series**, other titles occur in only four (*Mrs, Dr*) or one (*agent, detective*) of the series. Stubbs & Barth (2003:68) note that in contrast to other content words

Table 5. Raw frequency and dispersion for the interactive use of titles

	Raw frequency	Dispersion (episodes)	Dispersion (series)
Mr	44	15/32	5/7
Mrs	46	10/32	4/7
Dr	48	12/32	4/7
agent	49	4/32	1/7
detective	1	1/32	1/7

"sub-technical vocabulary […] occurs in predictable percentages, not only across sub-corpora of academic writing, but also in individual academic texts". While key words such as *Dr*, *agent* or *detective* are not exactly "technical", it is likely that they occur across series belonging to the same genre (e.g. medical or crime), but we would need to include more series of the same genre in the corpus to confirm this hypothesis.

Generally, the use of titles will reflect power hierarchies between characters — while informal language will be used by friends and equals (characteristic of the relationships in many sitcoms), formal titles will be used by strangers and unequals including those working in professional hierarchies (characteristic of many medical and crime series). For example, the two series classified as comedy above (*How I Met Your Mother, My Name is Earl*) only feature two interactive occurrences of *Mr* (both in *How I Met your Mother* — an exchange between a main character and a minor character), and zero interactive uses of *Mrs*, *Dr* or *agent*. This means that certain kinds of television dialogue will be *more informal* than unscripted spoken American English, while other kinds of television dialogue will be *more formal*. The hypothesis that television dialogue in general is more "informal" than unscripted spoken English can thus not be confirmed. My conclusion is that the

1 those shoulders, Commander. BOBBY: Agent Baer, Agent Todd, would either of
2 , Commander. BOBBY: Agent Baer, Agent Todd, would either of you like
3 problem, sir? BAER: Expect problems, Agent Todd. And with a little luck, you'll
4 One. KATE: It's not Air Force One, Agent Fornell. When the President
5 : Don't get into this pissing contest, Agent Todd. As you pointed out, the
6 long it took him to get here. ELMO: Uh, Agent Fornell here is FBI. Agent Todd,

Figure 3. Example concordances for *agent*

1 to go. Right now. SAM: McCreedy. Detective McCreedy, badge number 15A.
2 police uniform and make me call her Detective Sipowicz. KATE: What the
3 Susan: He dropped his gun, cleaning it. Detective Copeland: Miss Mayer, with all
4 your local post office. PAUL: Meaning? DETECTIVE: Meaning that the

Figure 4. Example concordances for *detective*

level of informality in TV dialogue depends on the genre of the television series as well as the specific character relationships (e.g. degree of power differences/hierarchies between characters) in a given series and episode.

3.3 How "emotional" is television dialogue?

In his analysis of *Friends*, Quaglio (2009) also argues that the language of this sitcom is more "emotional" than natural conversation, as evidenced by higher frequencies for intensifiers, certain discourse markers (*oh, wow*), stance markers (*of course*), non-minimal responses (*sure, fine*), expletives, the lexical bundles *I can't believe* and *Thank you so much*, emphatic *do*, the copular verbs *look, feel, sound* and certain slang terms (*suck, pissed off, screwed up, freak out*). These features are categorized by Quaglio as emotional language because they are resources that speakers use to "convey their feelings and express stance" (Quaglio 2009:87). Similarly, Culpeper (2001) talks about 'surge features' (ibid.:190) that are associated with affect and used by authors "as a conventional way of signalling that a character has a particular emotion or attitude" (ibid.:193). Drawing on Culpeper (2001), Quaglio (2009) and other research into language and emotion (Bednarek 2008, 2010), Table 6 lists selected features of emotional language, with a focus on affective stance — that aspect of stance concerning attitudes and desires (rather than, say, the epistemic or evidential status of what is being talked about — see Biber et al. 1999:972–975).

Table 6. Selected features of emotional language

evaluative language	*great, cool, suck, sure, fine, fortunately, I can't believe…*
emotion lexis	*happy, like, love, freak out, surprise, pissed off…*
modal verbs	*want, ought to, should…*
lexical repetition	*no no no no, hey hey hey…*
intensification	*really, so, totally…*
exclamations/interjections	*oh my god! wow!…*
expletives and taboo words	*bloody, hell, damn…*
emphasis	emphatic *do*
expressive speech acts	e.g. apologising, accusing, congratulating, complimenting, complaining, praising, boasting…

So is television dialogue in general (rather than just dialogue in the sitcom *Friends*) more emotional than unscripted conversation? To investigate this further, we need to consider which of the key words and trigrams in Table 2 can be associated with such emotional language. Here, the analyses clearly show that the

television corpus is indeed more "emotional" than the spoken corpus. In fact, with the exception of the semi-modals, informal vocabulary, pronouns and titles, all of the over-represented word forms and trigrams in Table 2 above can be associated in some way with emotionality. This has already been discussed with respect to the repeat *no no no*. Considering Table 6 above, it is also obvious that the expletive *hell* (and the trigrams *what the hell, the hell is*) and trigrams such as *oh my god*, *I can't believe*, *I love you* and *are you okay* are used as conventional ways of signalling different character emotions ranging from surprise to love and concern. Further, as I have argued above, even routinised expressions such as *thank*s or *sorry* are simply conventionalised expression of the emotions of gratitude and regret.

However, the extent to which key words/trigrams such as *hey*, *man*, *need to talk*, *out of here*, *I told you, I want you, you want me, want me to, why, what are you, why don't you, who are you, why are you, are you doing,* are also tied to the expression of emotion and stance needs further analysis.[5]

3.3.1 Hey *and* man

While *hey* in all its usages and *man* when used as vocative can be associated with informality (as an informal or colloquial resource of language), in some of their usages they can additionally be tied to emotionality. For example, many vocative uses of *man* occur in the immediate co-text of other features that point to emotionality, such as *oh man*; *I'm sorry, man* or in connection with an imperative (Figure 5, page 50). In such usages at least, I would argue that the vocative *man* can be very much associated with both informality *and* emotionality. If we also briefly consider *hey*, it is used in the corpus both as a simple greeting/attention getter (*Sarah, hey. It's Sam*) and, at other times, to additionally convey emotional meanings such as excitement, alarm or protest. The latter usage is illustrated in Figure 6 on page 51. As in Figures 1 and 2 above, the squares around particular parts of the concordances point to features that allow us to tie *hey* to the expression of emotionality. This includes interjections such as *my god* (cf. line 7 in Figure 6), exclamation marks and imperatives such as *easy, doc!* (l. 14), *step aside, man!* (l. 13), *wait!* (l. 11), expletives like *hell* (l. 10), evaluative language such as *cool!* (l. 3), demands such as *what're you doing?* (l. 4), *what's your problem pal* (l. 17), and accusations like *you're killing her* (l. 19). Again, in such usage, *hey* arguably indicates both informality *and* emotionality. In fact, Quaglio (2009) mentions a number of linguistic features that "perform multiple functions in the same context" (Quaglio 2009:143), and Culpeper & Kytö talk about the fact that "multiple functionality is typical" (Culpeper & Kytö 2010:147) for certain linguistic features.

3.3.2 Need to talk

The trigram *need to talk* can also be associated with matters involving emotions, most often to do with urgency. It is only rarely used in the corpus when referring to non-present participants (*We'll need to talk to him; I need to talk to him*) and most often seems to be concerned with matters of urgency between speakers and hearers, things that need to be "talked about". The following patterns occur in the corpus:

- *You need to talk to me/somebody*
- *We need to talk about X*
- *We (definitely/really) need to talk*
- *I need to talk to you in private/about X*

Such usages, I would argue, are tied to the expression of emotion in that they express emphasis through the use of the modal verb *need* and also imply that there is a serious or urgent matter that needs to be discussed. They are clearly not used neutrally but can be categorised under the general umbrella of "emotional language" as defined above.

3.3.3 Out of here

While the trigram *out of here* at first glance simply describes a location, it is in fact predominantly used in contexts where characters want to escape a situation they do not like. Consider Figure 7 on page 52 and note in particular the co-occurrence of *out of here* with other indicators of emotion (e.g. *hell*, exclamation mark) as well as expressions that imply that *getting out of here* is what is desired by speakers (e.g. *so I can…, come on, let's…, we'd all like to…, get us/me…,our ticket…*). The potential danger of the situation that characters want to escape is made explicit in concordance line 11: ***If** we make it out of here **alive**.*

1	on our fries. Darnell:cool. Earl: oh man. Randy: what's the matter. Earl: the
2	. Cuddy: Wear the coat. House: Man oh man. Someone got spanked real good
3	old gal, I swear. It's not funny. SAM: Oh, man, you should've seen your face, Dean.
4	a nazi cross in my yard. Scott: oh man. Earl: I got it. Scott; I'm burning my
5	Bandit. It's Bear-Awesome. Randy; oh man it is bear awesome. We're never
6	a mean round house kick . Marshall: Oh, man. Lily: Oh, well you did the right thing.

1	every done me. Marshall: Oh, come on, man, you don't mean that. Stuart: Yeah, I do.
2	I'll take it easy, all right! HURLEY: Hang on, man! You've got to slow down. MICHAEL: I'll
3	done something. DEAN: Come on, man, you risked your life. I mean, yeah,
4	: Where? DEAN: Vegas. What? Come on, man. Craps table. We'd clean up.
5	of things. What's it called? Earl: oh come on man! I'm just having a little trouble working
6	more intense. And painful. DEAN: Come on, man, it'll be alright. You'll be fine. SAM:

Figure 5. Example concordances for *oh man* and imperative + *man*

1 : Please! Hey, Roger, we're tryin' to help! Please! Hey! Hey, hey, hey, hey, hey, hey! ROGER: I don't
2 with you? SAM: Yeah-my head! DEAN: Sam? Hey. Hey! What's goin' on? Talk to me. SAM: It's
3 Puzzle bodies, a DA named Michael Grant. TONY: Hey, cool! After six months, we finally got a lead,
4 , under. BIKER: Hey! What're you doing? BIKER: Hey! Hey, hey! BIKER: My bike was stolen. You
5 going. TONY: Get up and run. Let's go! TONY: Hey! I just killed a guy back there. I am not getting
6 letting us clean toilets, other jobs like that. GIBBS: Hey, McGee, calm down MCGEE: Okay. Sorry.
7 . KATE: Hey! TONY: Your little quivering lip! KATE: Hey! My God, I swear the two of you are worse than
8 check out, you call me, okay? FOOTBALL KID: Hey, man, throw it! JEREMY DAVISON: it was ...
9 of giving it. TONY: Watch your lip, probie. KATE: Hey! TONY: Your little quivering lip! KATE: Hey! MY
10 ding. Oh, see you at practice. Sara: Ok. Reyes: Hey! where the hell are you going? Gibbs? Oh, you
11 , why don't you try it? GIBBS: Oh, wait a minute! Hey, wait! Wait a minute! Whoa! Stop! In here, in
12 SUN: Oh! MICHAEL: Hey, hey, hey! Hey, hey, hey, hey! MICHAEL: Hey, get your hands off her!
13 bumed my raft? JACK: Just calm down. MICHAEL: Hey, step aside, man! JACK: Just calm down! Calm
14 JACK: Michael! MICHAEL: Say it! SAWYER: Hey, easy, doc! One fight at a time. SAYID: He's
15 Lynette: Hey hey hey. HEY! Why are you stealing my garbage cans? Mrs
16 idea of helping is to let Tim just grope you?! Tim: Hey, you shouldn't talk to your mother like that,
17 this? Lamont: Yeah, mind you own business Tim: Hey, what's your problem pal? Lamont: You were
18 Don't thank me. It's just who I am. Foreman's voice: Hey! Foreman: Up and at 'em, big guy. House: Sorry
19 her up. That was Vogler, surgery's of. House: Hey! You're killing her! Vogler: Really? See, I
20 water? Next time, just take the whole bottle. Hank: Hey man, you got no right - House: See, people

Figure 6. Example concordances for *hey*

3.3.4 I told you

As Figure 8 on page 53 indicates, the trigram *I told you* is also often used in emotionally charged situations. For example, some of the occurrences in Figure 8 (e.g. lines 13–15, 19–20) seem to be instances where characters are reproached for not following orders, others seem to be instances where characters gloat/crow (e.g. lines 3, 11). In general, it is reasonable to assume that speakers telling listeners that they had told them to do something, or that something was the case or that something was going to happen is not a neutral speech act but is interpersonally loaded.

3.3.5 *Trigrams with* want *(*I want you, want me to, you want me*)*

Trigrams with *want* also deserve attention regarding their use in the corpus and their potential connection to emotionality. Table 7 shows the dispersion of the three trigrams, indicating that they are reasonably dispersed across different television series if not across all episodes.

Table 7. Frequency and dispersion of trigrams with *want*

	Raw frequency	Dispersion (episodes)	Dispersion (series)
I want you	18	13/32	7/7
want me to	18	14/32	6/7
you want me	15	12/32	6/7

In the TV corpus these trigrams with *want* (*I want you, you want me, want me to*) have to do with characters' wishes concerning other characters. As can be seen in Figure 9 on page 54, *I want you* is used primarily in the form *I* (=speaker) *want you* (= listener) *to do something*, but also includes a few occurrences such as *I want you out of the house!* (l. 4); *I want you out* (l. 3); *how do you know I want you in my dream* (l. 2).

Similarly, *you want me* is used by characters to enquire what other characters want of them. The majority of *you want me* is used in the forms *do you want me to do something*, or *what do you want me to do*, but also features the occurrences *You want me back*?; *You want me outta here*. With *want me to*, the majority again concern speakers and hearers (*you want me to; do you want me to; what do you want me to; if you don't want me to; if you want me to*), with only two occurrences including third persons (*How does he want me to fix it?; They want me to anchor the news tonight*.). Arguably, the majority of trigrams with *want*, then, clearly express attitudinal meaning in terms of the desires of speaker and hearer.

3.3.6 *Questions (*what are you, are you doing, who are you, why, why don't you, why are you*)*

Finally, let us briefly consider those key words and trigrams that seem to be associated with questions. This includes those with a *wh*-word (*why/what/who*) as

1 Me and you. We'll get Dean and Alice out of here. DEAN: Nuh-uh. No way.
2 this raft finished so I can get my boy out of here and take him home. SUN:
3 hurts. Earl: I used to sneak my car out of here all the time. They only turn
4 WHITE: And then we're gonna get out of here. you ready? TONY: I want
5 You okay? INMATE: Come on, let's get out of here! GIBBS: How d it go?
6 it is. It's a ballonnasaurus rex. Now get out of here. Mr Patrick: Hickey! You take
7 of-Woman: Uh, we'd all like to get out of here. Is there a problem? Lynette:
8 pilot your credentials, get us the hell out of here. GIBBS: Hey! Cockpit's on
9 it. So why don't you just get the hell out of here before I call the authorities
10 been kind of stressful. Willie: Pssh. I'm out of here. It's an expression. Bed.
11 GAVIN: Yeah? KAT: If we make it out of here alive. .we are so breaking up.
12 I'm thinking about transferring Matt out of here. Cameron: Ms. Davis, your
13 Help me! GAVIN: Kat! KAT: Get me out of here! SAM: Kat, it's not gonna hurt
14 FRED: Mary. MARY: Get these people out of here! FRED: They're federal
15 yeah, baby! TONY: That's our ticket out of here! All right, let's find out if
16 is brushing his tongue. Kate: Tony! Out of here! Now! Tony: What? I'm just
17 let's recap later, all right? Just get us out of here. DEAN: Well, I'm tryin' to pick
18 really grownup shades Gibbs: We're out of here. Willie: Thanks. Sweet. Ohh!
19 Kat. Come on, Sam's gonna get you out of here, and then we're gonna find

Figure 7. Example concordances for *out of here*

1 sound surprised, Agent Reyes. Reyes: I told you he was a suspect. Gibbs:
2 to help with the math test? Matt? Mom: I told you, he doesn't take drugs. Chase:
3 No... that won't be necessary. Mark: I told you! House: But I am going to
4 Yeah. DEAN: Yeah, exactly. SAM: I told you, I looked everywhere. I didn't
5 my husband in Iraq. LAURA ROWANS: I told you, I never saw that man before in
6 brought in. Cut to Reyes Reyes: Look, I told you, I can't. Not yet. We need his
7 or near the property. DEAN: Hey, man, I told you I searched that house up and
8 poodles nad funny hats right now. Earl: I told you. I don't know how to make
9 weren't clothes everywhere. Randy: I told you I'd help pick up after I watch
10 Would it make you feel any better if I told you I'm sorry for what I did? Bree:
11 left, and up the stairs. DEAN: See? I told you it'd work. SAM: Follow me. It's
12 ESP thing you've got goin' on. SAM: I told you, it's not ESP. I just have
13 J-just for a second. I'm sorry. JOHN: I told you not to leave this room. I told
14 JOHN: I told you not to leave this room. I told you not to let him out of your sight!
15 going out with a billionaire. Robin: Lily, I told you not to call him that. Ted: Wait,
16 restaurant, I still don't believe it. Lily: I told you. OK, a toast. Lily: Life is full of
17 idiots with sharp sticks. JACK: What if I told you that I had a way to get the
18 up new offices from Seattle to L.A. I, I told you that. Lynette: No. No, you
19 too stupid to bend balloons. Hickey! I told you to apologise. What's your
20 Darnell, what are you doing in here, I told you to wait in the car. Darnell: its
21 Sam, what's happening? SAM: I told you, you shouldn't have come.

Figure 8. Concordances for *I told you*

well as the trigram *are you doing* where the subject-verb inversion indicates use in a question. While *wh*-questions have been been grouped as features marking "involved" interactive texts (Biber 1988, Quaglio 2009:61) or classified as textual rather than interpersonal (Culpeper & Kytö 2010:131), Bubel (2006) and Bednarek (2010) have noted that in television series (*Sex and the City, Gilmore Girls*) questions can be concerned with the construal of affective relationships such as friendship, hostility, and conflict. Before discussing their functions in the TV corpus, let us again consider their dispersion (Table 8).

Table 8. Frequency and dispersion of key words/trigrams pointing to questions

	Raw frequency	Dispersion (episodes)	Dispersion (series)
why	299	32/32	7/7
what are you	59	24/32	7/7
are you doing	40	21/32	7/7
why don't you	29	22/32	6/7
who are you	14	9/32	5/7
why are you	14	10/32	5/7

1	asking myself that same question, Agent Gibbs. GIBBS: I want you and your Marines to check every car within a
2	know. Just to see what you're doing Earl: how do you know I want you in my dream Randy. I mean I could be with a lady
3	. Tess: I'm breaking up with you. When I get back home I want you out. Scott: ok honey. Have fun. could you grab
4	not yell at each other? I'm feeling really lousy today. Bree: I want you out of the house! Rex: Bree, look at me. It's not
5	be doing this now. Lets get you home Noah: There's a cop I want you to talk to… Mike: A cop? Noah: His name is
6	. GIBBS: Hey, McGee. MCGEE: Yes, boss? GIBBS: I want you to run the phone records for the land line at the
7	. ABBY: Or it just means he was never caught. GIBBS: I want you to run the DNA from his blood. Cross-reference
8	me up at nine. And don't be late Bree: Andrew, come here I want you to see this Andrew: What are you looking at?
9	that! Carly: What do you want me to do? Cry? House: Yes! I want you to tell me that your life is important to you,
10	: Charlie, I want her back. CHARLIE: What? I -- ETHAN: I want you to bring her back. CHARLIE: What did you do to
11	pricier, way more dangerous. On the other hand, it is legal. I want you to go to his house and find his stash. Betcha
12	MICHAEL: It's my job to look after him. JOANNA: Michael. I want you to turn on the "No Vacancy" sign while I'm gone.
13	DEAN: It's gonna be a lot louder than in the movies. So, I want you to stay under the bed, cover your ears, and do
14	. Ted: So, what? You want me to do a shot. Barney: Oh no. I want you to do five shots. Marshall: Oooh, more interesting.
15	one for your eyebrows. Earl: I feel fine sir. Mr Patrick: good. I want you to go in the back and sponge down the baby
16	: And then we're gonna get out of here. You ready? TONY: I want you to listen to me, Jeffrey. WHITE: Okay. TONY: I'm
17	gonna love it. Funsten: Why would I want to sue you? I want you to treat me. House: You're from Maplewood, New
18	. Hickey! I told you to apologise. What's your problem? I want you to apologise to these kids right now. Are you too

Figure 9. Concordances for *I want you*

As becomes apparent, all show a reasonable dispersion occurring in at least five of the seven different TV series in the corpus. The key word *why* and the key trigrams *what are you* and *are you doing* show the highest dispersion across individual episodes, as well as being most frequent in terms of raw frequency. (*Why don't you* has a slightly higher dispersion across individual episodes than *are you doing* but occurs less frequently.)

First, let us consider *what are you* and *are you doing*. Bednarek (2010) has shown how questions involving the n-gram *what … are you doing* are used in *Gilmore Girls* to express surprise or negative evaluation, but can also be employed as a more neutral inquiry. These different usages also occur in my data here, e.g.

1. Surprise: *What are you doing here?*
2. Negative evaluation/emotion: *What are you doing?!*
3. Inquiry: *What are you doing for Halloween?*

The first two are instances of an "emotional" usage. While surprise is perhaps less negative, the "demands for explanation" (Spitz 2005:316) of the form *What are you doing?!* are speech acts that belong to speakers' "argumentative resources" (Spitz 2005:245) and can thus be associated with argument or conflict between characters.

Considering the trigram *are you doing* in detail, out of 40 instances, the majority (35) are indeed instances of *what … are you doing*, which can be used to express negative evaluation/emotion and surprise as indicated above. Two further

instances, examples of the phrase *Why are you doing this?,* also clearly work as "demands for explanation" and can be tied to emotionality as well as conflict:

(4) HOUSE: Oh, will you stop it with the book! **Why are you doing this?**
CAMERON: I'm not doing anything.
HOUSE: You're manipulating everyone.
(*House*, 1.14)

(5) HOUSE: … It is not the hospital's fault if my son kicks off."
MOM: "Kicks off"?
HOUSE: I punched up the language; mostly for clarification. "I understand my doctors consider my decision to be completely idiotic-"
MOM: **Why are you doing this?**
HOUSE: "-but I am convinced that I know more than they do. I took a biology course in high school, so… yeah. Besides, I enjoy controlling every single aspect of my son's life, even if it means his death." Sign here please. I brought a pen.
MOM: Who are you?
(*House*, 1.08)

In (4), the demand for explanation is used by the main character (House, the leader of a special medical team in a hospital) to question Cameron's (a female member of his team) behaviour, with his negative emotionality also signalled in his subsequent accusation *You're manipulating everyone*. In (5), the second example from *House*, the demand is employed by the character of the mother to express her negative emotions towards House's inappropriate linguistic behaviour and to confront him about it. This example also shows that the trigram ***who** are you* can be used confrontationally, again as a "demand for explanation". This also seems to be the case in example (6), where it might additionally express confusion.

(6) MS. MILLER: Max, what's happening?!
MAX: Shut up!
MS. MILLER: What are you doing?!
MAX: I said shut up!
SAM: Max, calm down!
MAX: **Who are you?!**
SAM: We just wanna talk to you.
(*Supernatural* 1.14)

A state of emotional confusion also seems to be signalled by *who are you* in example (7) (from *Lost*). This dialogue exchange occurs when one of the female characters, Claire, wakes up in the presence of several other characters (including the

main characters Jack and Charlie), after having been abducted and clearly having lost her memory of prior events. The trigram and its repetition are used to signal her confusion and emotional anxiety.

(7) [Claire screams]
JACK: It's okay.
JACK: Claire. Hey, it's okay. It's okay.
CHARLIE: Claire. You're okay. You're safe.
CLAIRE: **Who are you? Who are you?! Who are you people?!**
(*Lost*, 1.15)

Let us now turn to the key word *why* to investigate its use as well as that of the key trigrams *why don't you, why are you*? Usages of *why don't you* are either friendly invitations (e.g. *why don't you join me for lunch*), suggestions (*why don't you check out those suspicious-looking containers, while you're at it*) or again confrontational (e.g. *why don't you show a little respect*). Usages of *why are you* similarly express either negative emotions tied to conflict (e.g. *why are you being this way*?), or less negative emotions such as surprise (e.g. *why are you sleeping in our tub*?).

It is beyond the scope of this paper to discuss all 299 occurrences of *why*, but the majority are indeed examples of direct questions (rather than other usages such as *I don't know why*/*that's why*, etc.) and can frequently be classified as demands for explanation and associated negative emotionality, particularly those involving what seems to be a direct accusation (e.g. *why can't you*, *why did you*, *why didn't you* in Figures 10 and 11).

To sum up, in this section I have argued that the key words and trigrams *hey*, *man*, *need to talk*, *out of here*, *I told you, I want you, you want me, want me to, why, what are you, why don't you, who are you, why are you, are you doing* are tied to the expression of emotionality, frequently in the context of negative emotions associated with conflict, confrontation or other problematic issues that characters face. Future research will need to investigate how such expressions of emotionality fulfil particular discursive functions in the specific contexts in which they occur, such as character revelation and control of viewer evaluation and emotions (Kozloff 2000, Bednarek 2010). However, we can clearly see that emotionality emerges as *the* key defining feature of the language of television.

This emotionality frequently seems to be negative and to be associated with conflict or drama. It may be a direct result of the professional practice of script writers, who are taught to write in ways which heighten dramatic conflict. For example, in a University of Technology, Sydney (UTS) workshop for budding scriptwriters on editing and improving their scenes (*Writing for Television*, 21/10/2008) convenor Tim Gooding, a scriptwriter himself, made many suggestions relating explicitly to the heightening of drama through the creation of more obstacles,

more resistance between characters, increased pressure, more conflict, more mystery, increased suspense, thus resulting in more impact on the audience. This is sometimes referred to with respect to a sense of urgency in writing handbooks: "A rule of thumb: *the greater the sense of urgency, the greater the dramatic conflict*. If a character desperately wants to achieve a goal, and some obstacle is thrown in the way, the dramatic tension heightens in direct proportion to that emotional intensity" (Blum 1995:82, italics in original). The creation of dramatic tension can in turn be tied to the aim of keeping audiences interested — after all, "the defining characteristics of film [and TV, M.B.] dialogue is that *it is never realistic; it is always designed 'for us'*" (Kozloff 2000:121, italics in original). Thus, this feature of television dialogue can be explained by reference to the "communicative context" of fictional television (Bednarek 2010), in particular with respect to its "double articulation" (Lorenzo-Dus 2009:161), where the language of television like other broadcast talk has to "meet the perceived demands of […] absent audiences, be it for information, entertainment, education, or a combination thereof" (Lorenzo-Dus 2009: see also Bubel 2006 on the notion of overhearer design). In other words, this emotionality is there for "us", the audience, with the aim of making us care and engage, and ultimately ensuring that we keep watching: "once people engage, they want to remain engaged" (Idato 2009:3), creating what Australia's Channel Ten chief programmer David Mott calls "stickiness" (in Idato 2009). This claim is also

1 have to be cola flavoured or cherry. Randy: Why can't you just give him back his wallet
2 Sophie: You know how blue I've been. Why can't you be supportive? Susan: I've
3 idiot for me, but I couldn't be Gretel. Why can't I be Gretel? Ted: Because you

Figure 10. Concordances for *why can't*

1 . I don't care what you think. Robin: Then why did you hide her from me? Ted: Wow,
2 . If it wasn't your intention to return, then why did you? SAYID: I was taken prisoner
3 even for one second. JACK: Charlie! JACK: Why did you do it, Charlie? CHARLIE:
4 didn't want to blackmail you. Gabrielle: Then why did you? Justin: I really needed to sleep
5 : Told me what? CLAIRE: Why?! CLAIRE: Why did you lie to me?! CHARLIE: Okay,
6 I ask you something? WALT: Sure. LOCKE: Why did you burn the raft, Walt? LOCKE:
7 was real. I threw up four times. Earl: wait why didn't you tell me about this. Randy:
8 Those bags are mine. DENNIS: Ohh! Why didn't you tell me you were subbing for
9 . JACK: You understood us all this time? Why didn't you say anything? SUN: Your
10 . And then I knew what I had to do. SAM: Why didn't you just leave? MAX: It wasn't
11 how you've chosen to deal with me? SUN: Why didn't you just tell me you didn't burn
12 here? Lynette: Uh, Mr Mullins had a spare. Why didn't you open the door? Mrs
13 check 'plus one' on the reply card. Lily: Why didn't you check 'plus... Ted: I did

Figure 11. Concordances for *why did/didn't you*

supported by Culpeper & Kytö's (2010:130) finding that *present-day play-texts*, as they call them, include a large amount of expressive n-grams, which they tie to the aim of these texts to create drama. It would be an interesting study in its own right to compare play-texts with television texts, as they are both scripted and both designed for audiences.

3.4 Which key words and trigrams are under-represented?

In the previous sections we have looked in detail at over-represented key words and trigrams, i.e. linguistic expressions that occur unusually frequently in the TV corpus. What about those key words and trigrams that are unusually **in**frequent? For ease of reference, Table 9 lists them again, and I have also tried to classify them into possible functional categories. The question marks signal that we would need to investigate each of these key words and trigrams in detail to ascertain whether or not they are indeed used in such functions in the specific contexts in which they occur.

Table 9. Negative key words and trigrams

	Negative key words and trigrams
vague language?	*kind, things; kind of a, it's kind of, something like that*
narrative-like language?	*had, and, then, they, them, they're; and it was, and then you*
discourse markers?	*well, know, think; I mean I, I mean you, you know I, that you know, and you know, you know you, you know the, you know it's, don't know I, yeah you know, I don't know, you know we, don't know if, you know it, but you know, I think that, and I think, yeah I think, I think that, I think it, I think the, I think it's, I don't think, yeah well I*
non-fluency features?	*uh, um, huh; uh you know, I I I, I I don't, uh I don't, uh you know*
intensifiers & quantifiers?	*so, really, a lot of, a little bit, some of the, one of the*
other	*but, because, it, I, that, that's, yeah, of, or, like, people, have, in; yeah yeah yeah*

Space does not permit a full discussion of these "negative" key words and trigrams, but, briefly, we can see that some of them can be associated with vagueness (Channell 1994), e.g. hedges like *kind of, it's kind of*; the vague coordination tag *something like that*; the noun of vague reference *things*. Others seem instances of narrative-like language, for example third-person pronouns, the trigrams *and it was*, *and then you* and the negative key words *then*, *had*. This ties in with findings by Mittmann (2006) and Quaglio (2009) that vague and narrative language is infrequent in television dialogue when compared to unscripted conversation. Other items in Table 9 can function as discourse markers (which Quaglio 2009:73

includes as vague language); three among those, *you know*, *I mean*, *I think,* are also under-represented in Mittmann's (2006) corpus. Quaglio (2009:73) also notes *you know* as more frequent in conversation but *I mean* as having a similar frequency as in *Friends*. Such discourse markers are very frequent in non-scripted conversation in general, functioning to negotiate interpersonal meanings as well as indicating vagueness or working as "fillers" in the discourse. This would also link them to non-fluency features, which I have already briefly discussed above and suggested as an area worthy of further investigation. Interestingly, some intensifiers and quantifiers are also under-represented, whereas a previous study (Tagliamonte & Roberts 2005) had found that in the sitcom *Friends* the same adjective intensifiers occur most frequently, as in non-scripted language (*really*, *very*, *so*). This led them to claim that "media language does reflect what is going on in language and may even pave the way for innovation" (Tagliamonte & Roberts 2005:280). In order to test this claim we would need to use concordances and concordance plots (showing dispersion within and across episodes) to compare the use of the key words and trigrams *so, really, a little bit, some of the, one of the* in the two corpora. This is because words such as *really* and *so* are multi-functional. For example, utterance-final *so* can indicate vagueness rather than intensification (Quaglio 2009:73) and *really* can be used in many different ways (Quaglio 2009:92), e.g. as intensifier or as interjection. This is also the case with many of the key words/trigrams listed in Table 9, so we need to use concordances and concordance plots to explore their dispersion, disambiguate between different uses and analyse the specific functions that they fulfil.

4. Conclusion

From the above analysis of key words and trigrams, emotionality emerges as *the* key defining feature of the language of television, cutting across individual series and different televisual genres. This can be linked to the creation of conflict and drama in televisual narratives. Is TV dialogue in general also more informal than unscripted conversation? The jury is still out on this matter. The level of informality featured in television dialogue seems to clearly depend on the genre, and perhaps even on the individual series. As the analysis of dispersion has shown, any linguistic analysis needs to be aware that there will always be linguistic features that do *not* occur across all televisual genres, series or even episodes. The language of a specific fictional television series will feature aspects related to the nature of fictional television series in general, but it will also include aspects particular to the genre it belongs to (e.g. we need to distinguish the language of sitcoms from the language of sci-fi (Mandala 2010)) as well as aspects unique to its character as

a particular popular cultural artefact (Bednarek 2011a).[6] That is, each series will have its own linguistic profile, being a cultural artefact in its own right — characterised by its own characters, settings, relationships, actions, events and language. In this sense we can talk about the language of *Friends* (Quaglio 2009), the language of *Gilmore Girls* (Bednarek 2010, 2011a) or the language of *The Wire* (Toolan 2011). Further research into television dialogue needs to be undertaken with the help of bigger corpora to investigate if a particular linguistic feature is characteristic of a series, of a particular televisual genre or of television dialogue as a language variety. While the assumptions made in this paper on the basis of a 130,000 word corpus of TV dialogue remain to be confirmed through the analysis of bigger corpora including dialogue from more series, I do suspect that the core features will remain stable. In other words, my claim is that there is in fact a fictional television register which might be less narrative and vague — and sometimes more informal, sometimes less informal — than unscripted language, but is best characterised by its high degree of emotionality.

Notes

1. For the comparison, the *p*-value was set to 0.000001, with a minimum frequency of 1 for key words and a minimum frequency of 2 for key trigrams. All the reported key words and trigrams are thus highly statistically significant. If the p value were increased, additional key words and trigrams would be found but they would be statistically less significant. It must be noted that even with this low *p*-value the software identified 792 key words (137 "negative", i.e. less frequent; 655 "positive", i.e. more frequent) and 271 key trigrams (39 "negative"; 232 "positive"). See Biber (2009) for an in-depth discussion of statistics in the retrieval of multi-word patterns.

2. Note that the corpus includes the names of the characters as speakers (e.g. *BOOTH: Bones!*) and dialogue only (no description of settings, camera action etc), mainly from the first or second season of a series, and that voice-overs and recaps were manually deleted; in the case of foreign-language dialogue (in *Lost*), this was deleted but the subtitled English dialogue was maintained. The fact that the transcripts need a considerable amount of (time-consuming) editing explains why the corpus used here is not bigger. Further edited transcripts will be added to the corpus in the future to represent additional series and genres.

3. Seventy-four out of 170 occurrences of *man* in my corpus are vocatives, and these occurrences are reasonably spread across series.

4. Semi-modals reflect "informality at the syntactic level" (Quaglio 2009:118).

5. Table 2 includes an additional key word indicating emotional language (*ok*), which is not discussed in this paper, as it might be a result of spelling variation (*ok* vs. *okay*).

6. A conceptualization of televisual genres in terms of a "prototype" or "family resemblance" structure may prove to be helpful. The notion of hybridity, or hybrid genres, would also need to

be incorporated, as contemporary "television plays fast and loose with genre boundaries, embracing the hybrid genre […] and self-reflexively referencing other genres" (Dunn 2005:138).

References

Baker P. 2005. *Public Discourses of Gay Men*. London/New York: Routledge.

Bednarek, M. 2008. *Emotion Talk Across Corpora*. Basingstoke/New York: Palgrave Macmillan.

Bednarek, M. 2010. *The Language of Fictional Television: Drama and Identity*. London/New York: Continuum.

Bednarek, M. 2011a. "The language of fictional television: A case study of the 'dramedy' *Gilmore Girls*". *English Text Construction*, 4 (1), 54–83.

Bednarek, M. 2011b. "The stability of the televisual character: A corpus stylistic case study". In R. Piazza, F. Rossi & M. Bednarek (Eds.), *Telecinematic Discourse: Approaches to the Language of Films and Television Series*. Amsterdam/Philadelphia: John Benjamins, 185–204.

Bednarek, M. 2011c. "Expressivity and televisual characterisation". *Language and Literature*, 20 (1), 3–21.

Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. 2009. "A corpus-driven approach to formulaic language in English". *International Journal of Corpus Linguistics*, 14 (3), 275–311.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Blum, R. A. 1995. *Television and Screen Writing. From Concept to Contract*. Boston: Focal Press.

Bubel, C. 2006. *The Linguistic Construction of Character Relations in TV Drama: Doing Friendship in* Sex and the City. Unpublished PhD dissertation, Universität des Saarlandes (Germany). Available at: http://scidok.sulb.uni-saarland.de/volltexte/2006/598/ (accessed December 2011).

Bubel, C. & Spitz, A. 2006. "'One of the last vestiges of gender bias': The characterization of women through the telling of dirty jokes in *Ally McBeal*". *Humor*, 19 (1), 71–104.

Channell, J. M. 1994. *Vague Language*. Oxford: Oxford University Press.

Culpeper, J. 2001. *Language and Charaterisation. People in Plays and Other Texts*. London: Longman.

Culpeper, J. 2009. "Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*". *International Journal of Corpus Linguistics*, 14 (1), 29–59.

Culpeper, J. & Kytö, M. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: Cambridge University Press.

Díaz-Cintas, J. (Ed.) 2009. *New Trends in Audiovisual Translation*. Clevedon: Multilingual Matters.

Dunn, A. 2005. "The genres of television". In H. Fulton, R. Huisman, J. Murphet & A. Dunn, *Narrative and Media*. Cambridge: CUP, 125–139.

Idato, M. 2009. "TV facing its own strict diet". *The Sydney Morning Herald*, *News Review*, May 2–3/2009, p.3.

Kozloff, S. 2000. *Overhearing Film Dialogue*. Ewing, NJ: University of California Press.

Lorenzo-Dus, M. 2009. *Television Discourse. Analysing Language in the Media*. Basingstoke/New York: Palgrave Macmillan.

Mandala, S. 2007. "Solidarity and the Scoobies: An analysis of the *–y* suffix in the television series *Buffy the Vampire Slayer*". *Language and Literature*, 16 (1), 53–73.

Mandala, S. 2010. *Language in Science Fiction and Fantasy. The Question of Style*. London/New York: Continuum.

Mittmann, B. 2004. *Mehrwort-Cluster in der englischen Alltagskonversation. Unterschiede zwischen britischem und amerikanischem gesprochenen Englisch als Indikatoren für den präfabrizierten Charakter der Sprache*. Tübingen: Gunter Narr.

Mittmann, B. 2006. "With a little help from *Friends* (and others): Lexico-pragmatic characteristics of original and dubbed film dialogue". In C. Houswitschka, G. Knappe & A. Müller (Eds.), *Anglistentag 2005, Bamberg–Proceedings*. Trier: WVT, 573–585.

O'Keeffe, A., McCarthy, M. & Carter, R. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Piazza, R., Bednarek, M. & Ross, F. (Eds.) 2011. *Telecinematic Discourse: Approaches to the Language of Films and Television Series.* Amsterdam/Philadelphia: John Benjamins.

Quaglio, P. 2008. "Television dialogue and natural conversation: Linguistic similarities and functional differences". In A. Ädel & R. Reppen (Eds.), *Corpora and Discourse. The Challenges of Different Settings*. Amsterdam/Philadelphia: John Benjamins, 189–210.

Quaglio, P. 2009. *Television Dialogue. The Sitcom* Friends *vs. Natural Conversation*. Amsterdam/Philadelphia: John Benjamins.

Rey, J. M. 2001. "Changing gender roles in popular culture: Dialogue in *Star Trek* episodes from 1966 to 1993". In D. Biber & S. Conrad (Eds.), *Variation in English: Multi-Dimensional Studies*. London: Longman, 138–156.

Richardson, K. 2010. *Television Dramatic Dialogue. A Sociolinguistic Study*. Oxford: Oxford University Press.

Scott, M. 2008. *Wordsmith Tools Version 5*. Liverpool: Lexical Analysis Software Ltd.

Scott, M. & Tribble, C. 2006. *Textual Patterns. Key Words and Corpus Analysis in Language Education*. Amsterdam/Philadelphia: John Benjamins.

Short, M. 1996. *Exploring the Language of Poems, Plays and Prose*. London: Longman.

Spitz, A. 2005. *Power Plays: The Representation of Mother-Daughter Disputes in Contemporary Plays by Women. A Study in Discourse Analysis*. Unpublished PhD dissertation, Universität des Saarlandes (Germany). Available at: http://scidok.sulb.uni-saarland.de/volltexte/2006/595/ (accessed December 2011).

Stokoe, E. 2008. "Dispreferred actions and other interactional breaches as devices for occasioning audience laughter in television 'sitcoms'". *Social Semiotics*, 18 (3), 289–307.

Stubbs, M. & Barth, I. 2003. "Using recurrent phrases as text-type discriminators: A quantitative method and some findings". *Functions of Language*, 10 (1), 61–104.

Tagliamonte, S. & Roberts, C. 2005. "So weird; so cool; so innovative: the use of intensifiers in the television series *Friends*". *American Speech*, 80 (3), 280–300.

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

Author's address

Monika Bednarek
University of Sydney, Australia
Department of Linguistics, F12
University of Sydney, N.S.W. 2006
Australia

Monika.Bednarek@usyd.edu.au

Copyright of International Journal of Corpus Linguistics is the property of John Benjamins Publishing Co. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.