Television dialogue and natural conversation

Linguistic similarities and functional differences

Paulo Quaglio
State University of New York at Cortland, USA

Motivated by ESL (English as a Second Language) concerns, this study compares the language of a U.S. situation comedy, *Friends*, with natural conversation. A corpus of transcripts of the television show and the American conversation subcorpus of the *Longman Grammar Corpus* are used for analysis. This data-driven investigation combines multidimensional (MD) methodology (Biber 1988) with a frequency-based analysis of a large number of linguistic features associated with the typical characteristics of face-to-face conversation. The results of the MD analysis indicate that *Friends* shares the core linguistic characteristics of face-to-face conversation, thus constituting a fairly accurate representation of natural conversation for ESL purposes. However, a closer look at the linguistic features revealed interesting functional differences between the two corpora. These differences pointed to distinct functional patterns (e.g., vagueness, emotional language) suggested by the association of linguistic features sharing similar discourse functions.

1. Introduction

There has been an increasing call for the use of authentic materials in the ESL (English as a Second Language) classroom (e.g., Burns, Gollin, & Joyce 1997; Carter & McCarthy 1994; Carter & McCarthy 1995; McCarthy 1998). In particular, the depiction of spoken language in ESL textbooks has been shown to be problematic. Research has revealed a discrepancy between the characteristics of naturally-occurring conversation and the dialogues found in ESL textbooks (e.g., Carter & McCarthy 1995; Koester 2002).

The discourse functions of conversation are linguistically realized by specific sets of grammatical features, which are rarely taken into account in ESL textbooks. As McCarthy and Carter (1995:211) put it, "speakers regularly make [grammatical] choices which reflect the interactive and interpersonal nature of the communication." To successfully interact in different social contexts, ESL students need

Corpora and Discourse

The challenges of different settings

Edited by Annelie Ädel and Randi Reppen

Studies in Corpus Linguistics

AMINS PUBLISHING COMPANY

to be aware of these grammatical choices and be exposed to the characteristics of spoken discourse (Burns, Gollin, & Joyce 1997). Spoken corpora can be instrumental in helping teachers and students accomplish this goal as "they offer direct access to characteristics of speech, so often inadequately described in textbooks" (Mauranen 2004: 89).

1.1 Television as a source of spoken data

Despite the overall agreement on the important role of natural conversation in ESL instruction, spoken corpora are not as readily available as written corpora (e.g., McCarthy 1998; O'Keeffe & Farr 2003). Teachers are then faced with the difficult task of collecting and transcribing such data themselves. A possible alternative to this time-consuming and costly solution is the use of television dialogue, which can be easily collected by teachers and brought into the classroom. Washburn (2001), for example, recommends the use of American situation comedies, especially for pragmatic language teaching and learning.

Even though the use of television dialogue as a surrogate for natural conversation for ESL purposes is appealing as well as practical, the language of television dialogue has not been fully analyzed from a grammatical point of view. Addressing the need to bring natural conversation into the ESL classroom (and the difficulty of obtaining spoken corpora), and recommendations for the use of situation comedies as a representation of face-to-face conversation, I take a corpus-based approach to compare the language of the U.S. sitcom *Friends* with naturally-occurring conversation. *Friends* was chosen not only because of its popularity but also because of its nature: a show about people who just sit around and talk, which makes it an interesting object of study for linguistic analysis, both as a comparison to natural conversation and as an object of study in itself.

2. Methodology

In this section, the two corpora used for analysis, *Friends* and the American English conversation subcorpus of the *Longman Grammar Corpus*, are described. A summary and examples of the major situational characteristics (types of settings and interactions) of each corpus are also included. Finally, an explanation of how the data were coded for analysis is provided.

2.1 The *Friends* corpus: Settings and interactions

The *Friends* corpus comprises *transcripts* (not scripts) of nine seasons of the show (from 1994 to 2003) and has approximately 600,000 words. The episodes were

transcribed and made available for entertainment purposes by several online fan clubs. The data used for analysis were taken from one of these fan clubs, *Crazy for Friends* (http://www.livesinabox.com/friends/). Transcripts of three episodes from each season (a total of 27 episodes) were randomly selected and compared with the actual videos of the shows and were considered fairly accurate and very detailed,

Table 1. Composition of the *Friends* corpus

Seasons	# of episodes	# of words	Average # of words/episode
1 (1994–1995)	24	60,180	2,507
2 (1995–1996)	23	65,364	2,842
3 (1996–1997)	25	67,994	2,720
4 (1997–1998)	23	71,732	3,119
5 (1998–1999)	24	57,460	2,394
6 (1999–2000)	23	69,652	3,028
7 (2000–2001)	23	60,882	2,647
8 (2001–2002)	23	76,205	3,313
9 (2002–2003)	24	75,298	3,137
Total	**206**	**604,767**	**2,935**

Table 2. Summary of settings and types of interactions in *Friends*

Settings	Types of interaction
Central Perk	Discussing things only guys can do (e.g., pee standing up) and only women can do (e.g., take out bra without taking off blouse) Discussing date plans for Saturday night Phoebe is coaching Chandler on how to break up with Janice Discussing plans for New Year's and how bad it is not to have a 'partner' Talking about how to quit smoking Playing the keyboard/friends listen and make comments
Monica and Rachel's apartment	Joey is trying to convince Monica to pose as his girlfriend Monica is trying to convince Rachel to waitress for her Monica & Phoebe are preparing for a barbecue for Rachel's birthday and talking about Joey's steady date
A fancy restaurant	A blind date situation
At a Laundromat	Ross is showing Rachel how to do laundry & 'hitting on' her
The ladies' room at a restaurant	Monica and Angela are talking about 'guys'
Chandler's office	Chandler interacts with supervisor
Monica's apartment	Making food
Chandler and Joey's	Discussing how Ross's date the previous night did not end up with sex
At the beach	Playing games; talking about sex partners; dating

including several features that scripts are not likely to present: hesitators, pauses, repeats, and contractions. Table 1 shows the composition of the *Friends* corpus.

A sampling of settings and types of interaction was carried out with the analysis of every fifth episode of each of the nine seasons (approximately 41 episodes). Table 2 shows a summary of the most frequent scenes and types of interactions identified in the *Friends* corpus. In spite of the presence of a few alternative settings (e.g., restaurant, Laundromat) throughout the show, most of the action takes place in two of the characters' apartments (Monica and Rachel's and Chandler and Joey's) and *Central Perk*, a coffeehouse where the group of friends meets regularly.

This brief analysis reveals not only a limited number of settings (as compared to the conversation corpus) but also an extremely restricted range of conversation topics, which typically involve relationships, love, dating, and sex. In Extract 1, the characters talk about their plans for New Year's. The topic of the conversation then shifts to relationships and how hard it is not to have a partner on such an occasion.

Extract 1: *Friends*, season 1, episode 10, *The one with the monkey*

Rachel: Hey, do you guys know what you're doing for New Year's? (They all protest and hit her with cushions) Gee, what?! What is wrong with New Year's?

Chandler: Nothing for you, you have Paolo. You don't have to face the horrible pressures of this holiday: desperate scramble to find anything with lips just so you can have someone to kiss when the ball drops!! Man, I'm talking loud!

Rachel: Well, for your information, Paolo is gonna be in Rome this New Year, so I'll be just as pathetic as the rest of you.

Phoebe: Yeah, you wish!

Chandler: It's just that I'm sick of being a victim of this Dick Clark holiday. I say this year, no dates, we make a pact. Just the six of us-dinner.

All: Yeah, okay. Alright.

Chandler: Y'know, I was hoping for a little more enthusiasm.

All: Woooo! Yeah!

2.2 The conversation corpus: Settings and interactions

The American English conversation subcorpus of the *Longman Grammar Corpus* has approximately 4 million words. Carefully designed to be representative of American conversation, it includes a wide range of settings (e.g., park, family home, classroom) and types of interaction (e.g., casual/task-related/telephone conversations). For the purposes of this study, a subcorpus of the American conversation corpus (of approximately 590,000 words) was utilized for analysis. This was done to make the searches for some of the linguistic features more manageable, as some of them had to be checked manually for disambiguation purposes.

Table 3. Composition of the American English conversation subcorpus

Speech Types	# of texts	# of words	Average # of words/text
Casual conversation	38	312,807	8,232
Task-related/service encounters/casual	19	152,819	8,043
Phone/casual conversations	16	108,347	6,771
Work-related only	2	15,749	7,874
Complete subcorpus	**75**	**589,722**	**7,863**

This sampling was proportional to the number of words in the four most frequent speech types (casual conversations, task-related and service encounters, phone and causal conversations, and work-related conversations) and settings and interactions described in the headings of each text.[1] The four speech types were included in the analysis (not only casual conversation) because the purpose of the study is to compare natural conversation in general to the language of *Friends*.

Table 3 shows the final composition of the American English conversation subcorpus utilized for linguistic comparisons. In addition to the number of texts that make up each of the speech type groups, the table shows the total number of words per group of speech types and the average number of words for each text in each of the four categories. These 75 texts containing a total of 589,722 words were utilized for analysis and will be referred to as the *conversation corpus*.

The overall analysis of settings and interactions within the four groups of speech types in the conversation corpus was based on the information contained in the file headers and qualitative analysis of several segments of dialogues. Table 4 gives a brief summary of the types of settings and interactions within each of the groups of speech types. This analysis reveals that each of the groups of speech types comprises a wide range of settings and interaction types/conversation topics.

Casual conversations refer to exchanges between family members and close friends; examples of task-related interactions include exchanges in a school in which students follow the teacher's instructions in a pottery class; service encounters include interactions at a community college's registrar's office (where students registering for classes interact with an attendant); the phone/casual conversation set of texts is characterized by casual conversations interrupted by occasional phone conversations, most of which show just one side of the conversation; finally, most of the work-related interactions occur in business offices. Extract 2 is an example of casual conversation, the most frequent speech type found in the conversation corpus.

1. For a detailed description of the sampling procedures, see Quaglio (2004).

Table 4. Summary of settings and types of interaction within speech types

Speech Type	Settings	Types of interaction/conversation topic
Casual Conversations	At home	Packing; playing games
	Home/bedroom	Chit-chat/gossip
	Café, parking lot	Women meeting at Starbuck's for coffee
	Condo/kitchen	Eating dinner and talking afterward
	Small restaurant	4 women meeting for dinner
	In the car	Chatting; talking about cultural issues
Task-related & Service Encounters	University building/ board room	Requesting money from Program Board for events
	Study room	Students studying, checking out books
	Faculty office	Discussion of collaborative article
	Small office space	Interacting with co-workers and supervisors
	Community college	In a pottery class
	Registrar's office	Registering for classes
	Dining room	Making Christmas cookies
Phone & Casual Conversation	Living room	Relaxing in the living room; chatting over the phone
	Restaurant/home	Having lunch with father; at home with mother; on the phone
	Kitchen/living room	Preparing dinner, eating, talking on phone
	Alterations shop	Casual conversation; phone call
	Private home	Family chit-chat; phone conversation
Work-related	Business office	Talk related to work activities
	House	Business meeting: discussing contents of a business letter

Extract 2. *Setting: Living room – Casual conversation*

A: So Larry did you manage to get any sleep beside Michelle's crying.
B: I didn't hear a thing.
A: Really.
B: Yeah.
C: God, I can't believe it.
B: I didn't hear a thing.

It is interesting to notice that most work-related interactions, as in Extract 3, tend to be rather casual in spite of the 'more serious' topics of the conversations. Notice, for example, the presence of incomplete sentences, a hesitator (*Uh*), a discourse marker (*well*), and a moderated expletive (Biber et al. 1999:1095) (*gosh*), which are typical features of face-to-face conversation.

Extract 3. *Setting: Business office – Work-related interaction*

A: Do you wanna attend the two-thirty meetings on Tuesday with the food vendors and the fire and healthy.
B: Uh yeah, I don't think I can. Well we're, let's say.
A: The College of Santa Fe.
B: College of Santa Fe gosh, I don't know that I can do that. . .well let me put it on the calendar, Dwight.

2.3 Data coding

The two corpora were annotated for parts of speech and various grammatical features using an automatic grammatical tagger developed by Douglas Biber. The *Biber Tagger* 'tags' texts for over 100 linguistic features. This grammatical annotation makes it possible to search for grammatical features or a combination of lexical items and grammatical features (e.g., *date* used as a noun – as opposed to the verb form). Below is an example of a tagged text; each tag is followed by its description.

I	^pp1a+pp1+++	[1st person personal pronoun]
have	^vb+hv+vrb++	[*Have* as main verb]
a	^at++++	[Indefinite article]
date	^nn++++	[Singular noun]
tonight	^nr+tm+++	[Time adverbial noun]

All of the searches for linguistic features (see the Appendix for a complete list) were done automatically using a concordance software program, MonoConc Pro 2.2 (Barlow 2002) and were manually checked for accuracy and disambiguation purposes.

3. Results

The present study combined multidimensional (MD) methodology (Biber 1988) with a frequency-based analysis of 166 linguistic features associated with the typical characteristics of naturally-occurring conversation. *Friends* was compared to conversation on Biber's Dimension 1 (*involved vs. informational production*), showing striking similarities. Despite these similarities, a closer look at the linguistic features revealed interesting functional differences between the two corpora. These functional differences were suggested by the association of linguistic features sharing similar discourse functions, which ultimately characterize each of the corpora.

In the following sections, I report on these results. First, I present the results of the MD analysis focusing on the similarities between the two corpora. I then describe the frequency-based comparisons of linguistic features and discuss the functional differences suggested by the results.

3.1 Multidimensional analysis: Similarities

Multidimensional (MD) analysis is a quantitative corpus-based technique designed to find and interpret the co-occurrence of certain linguistic features in a corpus. "On the assumption that co-occurrence reflects shared functions, analysts interpret the co-occurrence patterns to assess the situational, social, and cognitive functions most widely shared by the linguistic features" (Biber et al. 2002: 14).

The present study applied Biber's 1988 model of register variation focusing on Dimension 1. Briefly, high frequencies of features such as private verbs (e.g., *think, realize*), *that*-deletions, contractions, present tense, first- and second-person pronouns, the pronoun *it*, and demonstrative pronouns tend to co-occur in involved registers (e.g., face-to-face conversation), reflecting shared context, interactiveness, and real-time production, as in Extract 4.[2] Conversely, registers like news reportage and academic prose (as in Extract 5) are characterized by the co-occurrence of features such as nouns, nominalizations, attributive adjectives, and prepositions, reflecting the predominantly informational focus of these registers. The relevant features are underlined in both extracts.

Extract 4. *The conversation corpus*

1. Ira: I realize why I don't watch these sitcoms.
2. Brian: Cause they're stupid.
3. Ira: Yeah.
4. Amy: *Friends* is great.
5. Ira: No. That show is stupid.
6. Brian: It was great last year. It was really good <unclear> last year.
7. Ira: I hate sitcoms.
8. Brian: You watch Seinfeld?
9. Ira: That's, no I haven't watched it in forever and a day, but that's one of the only ones I actually <unclear>. Intelligent, humorous.

Extract 5. *Academic prose, Longman Grammar Corpus*

It is well recognized that a successful germination and establishment is one of the main contributing factors governing high yields.

---

2. See Biber, Conrad, and Reppen (1998), page 148, for a complete list of Dimension 1 features.

Table 5. *Friends* on Biber's (1988) Dimension 1

Scores	Selected Registers
35	*Face-to-face conversations*
34	*Friends*
30	
25	
20	*Personal letters*
15	
10	
5	*Prepared speeches*
0	*General fiction*
–5	
–10	*Press editorials*
–15	*Academic prose*

Once the *Friends* corpus was grammatically tagged, it was run through *Tag-Count*, a program developed by Douglas Biber. In simplified terms, this program counts the grammatical tags for each of the texts and outputs scores on each of Biber's dimensions of register variation by comparing these texts to those originally used in Biber's (1988) study.[3] For example, if a text receives a high positive score on Dimension 1, we conclude that it has a high frequency of the linguistic features characterizing involved registers, such as face-to-face conversation.

The MD analysis revealed a striking similarity between *Friends* and face-to-face conversation. Table 5 shows that the score obtained by *Friends* (34.4) on Dimension 1 was very similar to that of conversation (35.3), indicating a high degree of involvement. Despite the similarities, it should be noted that the Standard Deviation for *Friends* (4.3) was much smaller than that of conversation (9.1), suggesting that the sitcom presents much less variation when compared to conversation. This difference seems to result from the much narrower range of speech types found in *Friends*, as revealed by the analysis of settings and interaction types discussed in Section 2.

Extract 6: *Friends*, season 2, episode 11, *The one with the lesbian wedding*

Joey: Pheebs, **who's** Evelyn Dermer? [contraction, pres tense vb, *wh*-question]
Phoebe: **I don't know.** [1st pers pron, contraction, private vb, pres tense] **Who's** Soupy Sales? [contraction, pres tense vb, *wh*-question]
Mrs. Green: Oh my god, **there's** an unattractive nude man playing the cello. [contraction, pres tense vb]

3. Refer to Biber (1988), Chapters 5 and 6, for a thorough description of multidimensional analysis and how the scores were computed.

Rachel: Yeah, **well just be** glad he's not playing a smaller instrument. [discourse particle, emphatic, *be* as main vb, pres tense vb, contraction]

Mrs. Green: **You have** some life here, sweetie. [2nd pers pron, pres tense vb]

Rachel: **I know.** [1st pers pron, private/pres tense vb] **And** Mom, **I realize you** and Daddy were upset when I didn't marry Barry and get the big house in the suburbs with all the security and *everything*, [non-phrasal coord, 1st/2nd pers pron, private/pres tense vb, *that*-deletion, 1st pers pron, indef pron] but **this is just so** much better for **me, you know?** [dem pron, *be* as main vb/pres tense, emphatics, 1st/2nd per prons, private/pres tense vb/discourse marker]

Mrs. Green: **I do.** [1st pers pron, *do* as pro vb, pres tense vb] **You** didn't love Barry. [2nd pers pron, contraction] **And I've** never seen **you** this happy. [Non-phrasal coord, 1st pers pron, contraction, 2nd pers pron] I look at **you** and **I think**, oh, **this is** what I want. [1st pers pron, pres tense vb, 2nd /1st pers prons, private/pres tense vb, dem pron, *be* as main vb, pres tense vb, 1st pers pron, pres tense vb]

Rachel: For...**me.** [1st pers pron]

Mrs. Green: **Well,** not **just** for **you.** [discourse particle, emphatic, 2nd pers pron]

Rachel: **Well, what do you mean?** [discourse particle, *wh*-question, 2nd per pron private/pres tense vb]

Among several other linguistic features, Extract 6 has 11 first-person pronouns, 8 second-person pronouns, 8 contractions, 16 present-tense verbs, and 6 private verbs. Most of this dialogue involves one of the main characters of the show (Rachel) and her mother (Mrs. Green), talking about personal issues – primarily the acknowledgement by Mrs. Green that her daughter's decision not to marry her fiancé was right. In addition, Mrs. Green's realization is woven with a tinge of envy of her daughter's happiness. Later in the segment, Mrs. Green reveals her plan to leave her husband, accentuating the personal nature of the exchange. This short segment illustrates how the linguistic features of Dimension 1 (in bold, descriptions of features in square brackets) co-occur in *Friends*, reflecting a high degree of involvement.

In conclusion, the results of the MD analysis indicate that *Friends* shares the core linguistic characteristics of face-to-face conversation, thus constituting a fairly accurate representation of natural conversation for ESL purposes. Extract 6 shows that the language of *Friends* closely resembles that of naturally-occurring conversation as revealed in Biber's (1988) study of register variation. Next, I compare *Friends* to conversation from a functional standpoint.

3.2 Functional analysis of *Friends*

Language is used for different purposes, such as to inform, to analyze, to describe, to persuade, and to express stance and personal feelings. These discourse functions are realized by sets of co-occurring linguistic features. The discourse circumstances of conversation (e.g., shared context, real-time production, interactiveness) (Biber et al. 1999: Ch. 14) are reflected by specific sets of linguistic features. For example, the shared context of conversation (location, knowledge) "is often reflected linguistically in the simplification of grammatical structures" (Quaglio & Biber 2006: 705). The apparent vagueness resulting from this simplification is reflected in the high frequency of features, such as first- and second-person pronouns, hedges (e.g., *kind of*), and nouns of vague reference (e.g., *stuff*).

In this section, I compare *Friends* to conversation relative to the association of linguistic features sharing similar discourse functions. I focus on two of these functions: vagueness and emotional content.[4]

3.2.1 *Vagueness*

Consonant with the MD analysis, most of the linguistic features associated with conversation revealed an overall high frequency in both corpora. A closer analysis, however, showed that certain sets of features tended to have much higher frequencies in one of the two corpora. These features were then grouped according to the discourse functions they shared. Table 6 shows that the vast majority of the features associated with vague language are more frequent in the conversation corpus. In this section, I focus on the description of a selection of these features: hedges, coordination tags, nouns of vague reference, and the discourse marker *you know*.

Hedges, coordination tags, and nouns of vague reference

Conversation tends to be vague due to the shared context and pressures of real-time production (Biber et al. 1999: Ch. 14). It makes extensive use of devices of vague reference, including hedges (e.g., *sort of, kind of*), coordination tags (e.g., *and stuff like that*), and nouns of vague reference (e.g., *stuff, thing*), which are probably the most obvious features associated with vagueness. Figure 1 groups together each of these devices of vague reference showing that each of these groups of features is more frequent in the conversation corpus and that the overall count of these vague devices (5,228 times/million words in conversation and

4. See Quaglio (2004) for a detailed analysis of the functional differences and multiple examples.

Table 6. Features associated with vague language

Categories	Feature	Conv	*Friends*	Both
Hedges	*kind of; sort of*	•		
Coordination tags	*or something (like that)*	•		
	or anything (like that)	•		
	(and) stuff (like that)	•		
Vague reference	*thing(s); shit*	•		
Discourse marker	*you know;*	•		
	I mean			•
Stance markers	*probably*	•		
	perhaps; maybe		•	
Modals	*could*			•
	might	•		
Copular verbs	*seem; appear*			•
Utterance final *so*	*so*	•		

3,745 times/million words in *Friends*) is almost 1.5 times more frequent in the conversation corpus.[5]

Examples (1) through (6) illustrate the use of these features.

(1) A: and, uh, he showed them to some university professors at UNM or someone did and now <unclear> he's got some sort of honorary degree.
B: I think they, they gave him permission to go fossil hunting in places where only university folks can go. (Conversation)

(2) Monica: I hate men! I hate men!
Phoebe: Oh no, don't hate, you don't want to put that out into the universe.
Monica: Is it me? Is it like I have some sort of beacon that only dogs and men with severe emotional problems can hear? (*Friends*)

(3) A: Presto Pasta it's a fast food pasta place.
B: Pasta's kinda heavy though. (Conversation)

(4) Rachel: Okay, Monica, y'know what, honey, you're kinda losing it here! I mean, this is really becoming like a weird obsession thing. (*Friends*)

(5) A: If you want to check out that shrine on Sunday or something
B: Yeah, that would be cool. (Conversation)

(6) Phoebe: You guys wanna try and catch a late movie or something?
Rachel: Maybe, but shouldn't we wait for Chandler? (*Friends*)

5. Even though the two corpora had a similar size, I report the results in occurrences per million words for ease of comparison with other studies, which tend to use large corpora and report the results in this manner.

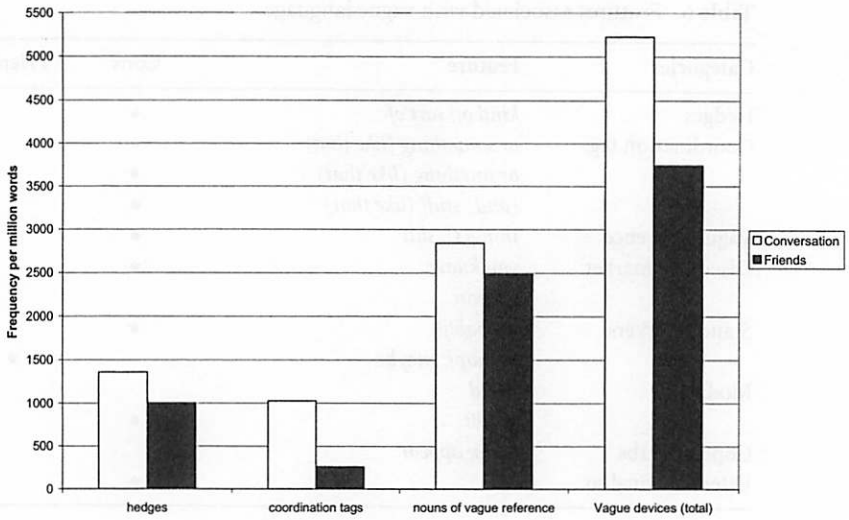

Figure 1. Frequency of hedges, coordination tags, and nouns of vague reference

The apparent imprecision caused by hedges has, in fact, important discourse functions. As Leech (2000: 695) puts it, hedging expressions "allow[ ] a speaker to take refuge in strategic imprecision." In (1) and (2), the speakers are aware that *honorary degree* and *beacon* are not the most precise terms to use. The hedge *sort of* is an acknowledgment of this lack of precision and can be an 'invitation' for the interlocutor to collaboratively construct the intended meaning. In other words, hedges "make it easier for the listener to pick out the specific referent the speaker has in mind if the linguistic expression is not exact" (Aijmer 1984: 122). Further, hedges create a desirable sense of vagueness, which may lead interlocutors to actively participate in the interaction by asking clarification questions and volunteering possible interpretations. This imprecision is ultimately a communicative device which facilitates the interaction between speakers, making it a dynamic process of verbal exchanges.

The hedge *kinda* in (3) and (4) reflects a non-confrontational tone in the exchanges. McCarthy and Carter (1997) suggest that the undesirable effect that overly direct utterances can create is functionally mitigated by the imprecision brought about by the use of hedges. In (3), this softening effect is enhanced by the use of the linking adverbial *though* at the end of the utterance. When used in final position in conversation, "*though* makes the disagreement much softer than a marker of direct contrast, such as *but* or *however*" (Biber et al. 2002: 394). In (4), the same soothing effect is created when *kinda* precedes the potentially rude or overly straightforward *losing it here.* Notice that Rachel's use of the term of endearment *honey* further cushions the potentially face-threatening utterance. The

coordination tag *or something* in (5) and (6) produces a similar mitigating effect in that it suggests flexibility and a desire not to impose.

Discourse marker you know

The discourse marker *you know* is 3 times more frequent in conversation: it occurs 4,990 times/million words in the conversation corpus and 1,563 times/million words in *Friends*. Aijmer (1984) suggests that because *you know* often collocates with *kind of* and *sort of*, it shares their hedging function, thus contributing to the vague nature of conversation. In (7), *you know* collocates with *kind of* and is preceded by *sort of* intensifying the overall vagueness of the utterance; in (8), *y'know* is preceded by *kinda* (and the downtoner *just*), which contribute to accentuate the mitigating effect intended by the speaker.

(7) A: you could just sort of open them up yeah and just you know kind of spread them out. (Conversation)

(8) Rachel: Hi! Sorry- sorry we're late, we, uh, kinda just, y'know, lost track of time. (*Friends*)

The higher frequency of the linguistic features discussed in this section suggests that conversation in general tends to be more vague than the conversation presented in *Friends*. Vagueness is less desirable in *Friends*, as the audience (the interlocutors of the show) cannot interact with the characters. It seems like the use of vague devices in *Friends* is constrained by a "clarity cut-off boundary," beyond which comprehension can be adversely affected. However, it is important to point out that all of the vague devices with their different intended functions do occur in *Friends* but to a much lesser extent.

3.2.2 *Emotional language*

In involved spoken registers, such as casual conversation, participants express feelings, attitudes, and concerns. This involvement is reflected in the speakers' tone of voice, intonation patterns, nonverbal signals, and linguistic features. In this section, I focus on the linguistic choices speakers make to convey their feelings and express stance. I use the cover term *emotional language* here to refer to any emphatic form of expression that is captured by the use of certain linguistic features and can thus be identified in a transcribed corpus and studied from a grammatical standpoint.

Several features have been associated with the expression of stance and emphatic content. Among these features are adverbial intensifiers (e.g., *really*), some inserts (e.g., *wow*), stance markers (e.g., *of course*), and expletives/taboo terms (e.g., *damn*). Based for the most part on a survey of the *Longman Grammar of Spoken and Written English* (Biber et al. 1999), 33 features associated with the expression of emotion and/or emphatic content were chosen for analysis. Table 7

Table 7. Features associated with emotional language and/or emphatic content

Category	Feature	Conv	*Friends*	Both
Intensifiers	*so; really; totally*		•	
	too			•
	damn			•
Inserts	*oh; wow*		•	
Stance marker	*of course*		•	
Non-minimal responses	*wow*	•		
	sure; fine		•	
Expletives	*damn; bastard; bitch(y)*		•	
	son of a bitch		•	
	shit(ty)	•		
	fuck [and variations]	•		
	suck; screw(ed)(up)		•	
	piss(ed)(off)	•		
	ass; crap(py)		•	
Innovations	*all* + adjective/gerund		•	
	so + verb		•	
	so (not) + NP		•	
	so not + Adj		•	
	totally [emphatic agreement]		•	
Lexical bundles	*I can't believe* (+complements)		•	
	thank you so much		•	
Emphatic do	*do*		•	
Copular verbs	*look; feel; sound*		•	

shows that 27 of these features were more frequent in *Friends*, two had similar counts, and only four of them were more common in conversation. Next, I describe and illustrate some of these features: adverbial intensifiers, expletives/taboo terms, and linguistic innovations.

Adverbial intensifiers so, really *and* totally

Among the more obvious features indicating emphatic and emotional content are adverbial intensifiers, such as *so*, *really*, and *totally*. Biber et al. (1999:564–6) report that amplifiers (adverbial intensifiers) are most common in conversation and that speakers use a wide range of informal intensifiers to express stance, emotion, and for emphatic purposes. *Really* occurs 3,456 times/million words in the conversation corpus and 3,968 times/million words in *Friends*. In (9), the repetition of *really* intensifies its amplifying effect; in *Friends*, this repetition occurs over 3 times more often than in conversation (108 versus 34 times/million words, respectively). It is also interesting to notice that the distribution of the repetition *really really* is more 'balanced' in *Friends*: it modifies either adjectives or adverbs 61.5% of the time and precedes verbs in 38.5% of the instances. The modification of ei-

ther adjectives or adverbs is preferred 75% of the time in the conversation corpus, as in Example (10), whereas verbs are modified in only 25% of the occurrences. Even though the difference in the overall frequency of *really* in *Friends* and conversation is not striking, this repetition phenomenon (along with the important differences described below) contributes much to the more emphatic/emotional nature of the language of *Friends*.

(9) Rachel: You really, really need to get some sleep, honey.
Monica: I know I do. (*Friends*)

(10) A: And I said, I can't do that and she got really, really angry with me. (Conversation)

So, as in (11), is 1.7 times more common in *Friends*, occurring 1,449 times/million words versus 842 times/million words in the conversation corpus. In addition to this higher frequency, the most common adjectives modified by *so* in *Friends* are *sorry* and *glad*; in the conversation corpus, the most frequent right collocates (adjectives) are *good* and *cute*, suggesting a more personal (or affective) use of this adverbial intensifier in *Friends*. *Totally* is over twice more frequent in *Friends*, occurring 402 times/million words and 180 times/million words in the conversation corpus. Notice that, in (12), *totally* modifies the adjective *tired* and is interchangeable with *completely* or *really*. A more recent use of *totally* is illustrated in (13): it is more commonly used by younger speakers of American English and has the meaning of *for sure*. Whether used in its more canonical sense or with its innovative connotation, *totally* reflects emphatic/emotional content and is much more frequent in *Friends*.

(11) Chandler: Oh man, I am so sorry. Are, are you okay?
Joey: Well, I've been better. But, I'm all right. So you like her huh? (*Friends*)

(12) A: I was like totally tired, you woke me up and I was home alone when you called to play basketball?
B: It's alright You're going today? (Conversation)

(13) Rachel: I was giving you an apology and you were totally checking her out! (*Friends*)

Expletives/taboo terms

Following Stenström (1991), I lump together expressions or words commonly referred to as expletives, taboo words, and swearwords in this analysis and refer to them as *expletives*. Expletives in general are strongly associated with the expression of emotion; they "are realized by taboo words related to religion, sex and the human body, which are used figuratively and express the speaker's (genuine or

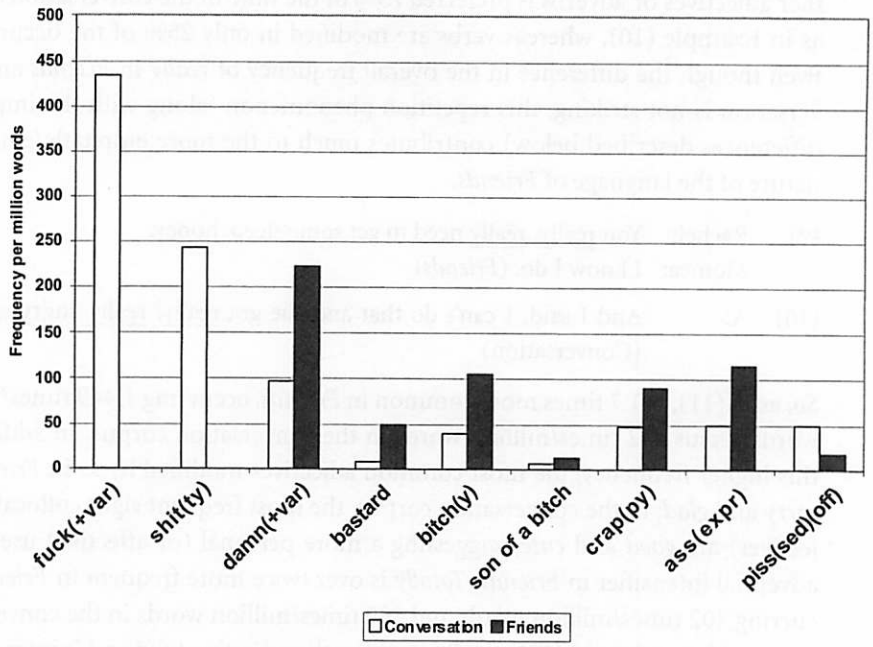

Figure 2. Frequency of expletives associated with emotional language

pretended) emotions and attitudes" (Stenström 1991:240). Figure 2 displays the frequency of the expletives selected for analysis. Surprisingly, except for *fuck*, *shit*, and *piss(ed)(off)*, all of these expletives are more frequent in *Friends*. I comment on and exemplify the use of most of these expletives below.

Fuck (plus variations) and *shit(ty)* are by far the most frequent expletives in conversation, occurring 435 and 244 times/million words, respectively. In (14), *fucking* is an adverbial intensifier, and *shit* an emotionally-loaded noun of vague reference;[6] both *shit* and *fuck* in (15) are used as exclamatory inserts and are thus instrumental in conveying speaker A's strong dissatisfaction with the fact that his computer file has been deleted. Because of restrictions and regulations imposed by the televised medium, these terms are not part of the *Friends* lexicon. Such limitations may be responsible for the overuse of the adverbial intensifiers discussed above. Further, the restrictions of the use of *fuck* as an exclamatory insert and *shit* as an exclamatory insert, a (emotionally-loaded) noun of vague reference, and as an evaluative adjective (*shitty*) seem to explain the overuse of *crap(py)* in *Friends*: it is twice more frequent than in the conversation corpus, occurring 90 times/million

6. I call this use of *shit* 'emotionally-loaded noun of vague reference' to differentiate it from the apparently neutral use of *shit* as a noun of vague reference, as in "... yeah there's some shit in here I would buy but I hate buying shit out of a catalog, personally" (Conversation).

words, as an apparent compensation strategy. In (16), it is used as an exclamation and in (17) as an emotionally-loaded noun of vague reference.

(14) A: I would kick someone's fucking ass, if anyone put shit like that on me ... here, drink these. (Conversation)

(15) A: It's not a memo any more there or something.
B: Oh shit, no wonder I, fuck, they wiped my Word out. (Conversation)

(16) Chandler: Ho, ho, ho, holy crap is it hot in here!
Joey: Really, hey, you mind if I turn the heat down? (*Friends*)

(17) Chandler: Y'know, of all my friends, no one knows the crap I go through with my mom more than you. (*Friends*)

The also surprisingly higher frequencies of *damn* (plus variations/combinations, such as *damnit* and *goddamn*), *bastard*, *bitch*, and *son of a bitch* in *Friends* may be the result of the same compensatory phenomenon discussed above. Even though the nature and grammatical uses of these terms are different from those of *fuck* and *shit*, in conjunction with other emotionally-loaded language they reflect the speakers' emotions and feelings more emphatically than other less harsh terms perhaps would. *Damn* is over twice more frequent in *Friends*, occurring 223 times/million words versus 97 times/million words in the conversation corpus); *bastard* occurs 49 times/million words in *Friends* and only 8 times/million words in the conversation corpus; altogether, *bitch(y)* and *son of a bitch* occur almost 3 times more often in *Friends* (127 times/million words). All taken from *Friends*, examples (18) through (21) exemplify the use of these expletives.

(18) Monica: Damnit Ross, get your butt out of the bathroom.
Ross: Calm down, I'm blow-drying. (*Friends*)

(19) Mrs. Geller: Well what is it? Come on sweetie, you're like, freaking me out here.
Ross: I hate Chandler, the bastard ruined my life. (*Friends*)

(20) Ross: Well ah, Aunt Silvia was, well not a nice person.
Monica: Oh, she was a cruel, cranky, old bitch! (*Friends*)

(21) Chandler: All right! Go left! Go left! Go right!! Go right!!
Phoebe: I can't!! I can't!! Noooooooo!!!!!!! You son of a bitch!!!!! (*Friends*)

Innovative uses of all, so, *and* totally

Linguistic innovations are often associated with the expression of stance (Quaglio & Biber 2006). I included the innovative uses of *all*, *so*, and *totally* here because of their emphatic characteristics and apparently increasing frequency in American

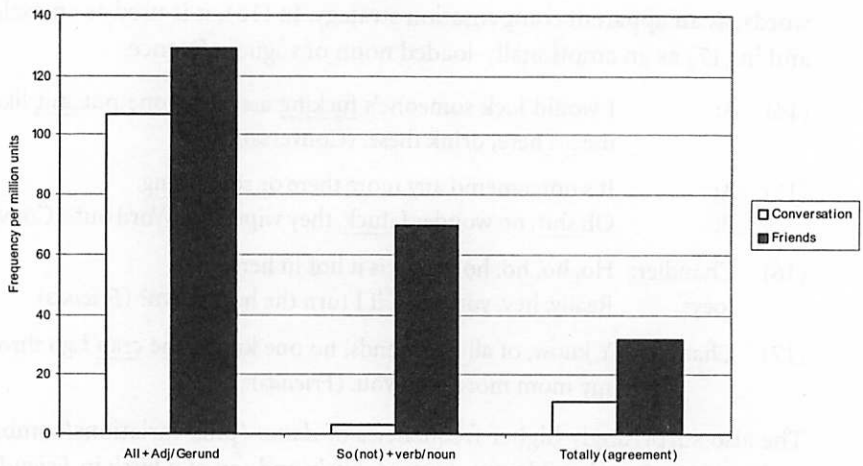

Figure 3. Frequency of linguistic innovations associated with emotional language

English conversation. Figure 3 shows that the frequencies of these three features are higher in *Friends*.

The relatively new use of *all* followed by an adjective or a gerund is described by Waksler (2001:128) as "a marker of the speaker's upcoming unique characterization of some entity in the discourse;" this characterization is typically emphatic in nature. In (22), *all* intensifies the adjective it precedes (*pissed*); in (23), its emphatic content seems to spread to the whole chunk of discourse following it. The adverbial intensifier *so* was discussed at the beginning of this section. As a linguistic innovation, *so*, which typically precedes an adjective, modifies a noun, as in (24), or a verb, as in (25). As an intensifier of nouns and verbs, *so* is 23 times more frequent in *Friends*, occurring 70 times/million words; in the conversation corpus, it occurs only 3 times/million words. This discrepancy is probably due to the fact that the conversation corpus was collected between 1995 and 1996. This innovation then might not have been fully captured by the conversation corpus.

(22) A: Yeah, let him deal with Chris <unclear>.
B: Okay. He'll be <u>all</u> pissed though. (Conversation)

(23) Chandler: What are you talking about?
Joey: She was <u>all</u> crying. She-she said you guys want different things, and that and that she needed time to think. (*Friends*)

(24) Ross: Please. This is <u>so</u> your fault.
Susan: How, how is this my fault? (*Friends*)

(25) Rachel: And even though I am <u>so</u> looking forward to the next part, I am really gonna miss being pregnant. (*Friends*)

Another fairly recent innovation in American English conversation is the use of *totally* not as an adverbial intensifier but as a self-contained expression of emphatic agreement, as in (26). With this function, *totally* is 3 times more frequent in *Friends*, occurring 32 times/million words; in the conversation corpus, it occurs 11 times/million words. It never occurs in the first turn of an interaction; rather, it is a response and typically occurs by itself. Semantically, it expresses more than simply "I agree with you"; it also shows stance in that it suggests unconditional agreement.

(26) Chandler: That's a great idea! We can easily think of a way for us both to enjoy the room.
Monica: Totally! (*Friends*)

This section has focused on the linguistic markers of emotionally-loaded language. The frequency-based analysis of these markers revealed that the language of *Friends* tends to be more emotional and emphatic than conversation. This dramatic effect is linguistically realized by a high frequency of several features associated with (but not limited to) emotionally-loaded language, such as adverbial intensifiers, expletives, and linguistic innovations. It is important to emphasize that all of the features found in the conversation corpus (except for the expletives *fuck* and *shit*) are also found in *Friends* and vice versa. What differentiates one corpus from the other is the frequency in which these features occur.[7]

4. Conclusion

The MD analysis of *Friends* showed that this sitcom shares the core linguistic features of conversation. A closer frequency-based analysis of a large number of linguistic features associated with conversation revealed interesting differences between *Friends* and conversation at the functional level; two of these differences were addressed in this paper: vagueness and emotional language.

On the one hand, *Friends* has a lower frequency of the linguistic features associated with vague language, such as hedges and nouns of vague reference; on the other hand, the language of *Friends* is much more emotionally-loaded, as evidenced by the much higher frequency of features such as adverbial intensifiers and expletives. These results may be a reflection of situational differences between *Friends* and the conversation corpus. The typical vagueness of conversation is undesirable in *Friends*, as it can ultimately lead to incomprehensibility. The higher

7. Evidently, there are several differences between television dialogue and natural conversation. For example, *Friends* has virtually no instances of overlap, interruptions, and unclear words. Unlike in conversation, the turns tend to be evenly distributed in *Friends*. These important differences, however, are beyond the scope of the present study.

emotional content of *Friends* can arguably be a reflection of the dramatic nature of television dialogue and/or a result of the close relationship shared by the characters. Despite such differences, addressing the original ESL-motivated research question that guided the present study, the language of *Friends*, overall, is a fairly accurate representation of face-to-face conversation.

References

Aijmer, K. 1984. *Sort of* and *kind of* in English conversation. *Studia Linguistica* 38: 118–128.

Barlow, M. 2002. *MonoConc Pro* (Version 2.2) [Computer software]. Houston TX: Athelstan.

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

Biber, D., Conrad, S. & Reppen R. 1998. *Corpus Linguistics: Investigating language structure and use*. Cambridge: CUP.

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

Biber, D., Conrad, S., Reppen, R., Byrd, P. & Helt, M. 2002. Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly* 36: 9–48.

Burns, A., Gollin, S. & Joyce, H. 1997. Authentic spoken texts in the language classroom. *Prospect* 12(2): 72–86.

Carter, R. & McCarthy, M. 1994. *Language as Discourse: Perspectives for language teaching*. London: Longman.

Carter, R. & McCarthy, M. 1995. Grammar and spoken language. *Applied Linguistics* 16(2): 141–158.

Koester, A. J. 2002. The performance of speech acts in workplace conversations and the teaching of communicative functions. *System* 30: 167–184.

Leech, G. 2000. Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning* 50(4): 675–724.

Mauranen, A. 2004. Spoken corpus for an ordinary learner. In *How to Use Corpora in Language Teaching*, J. M. Sinclair (ed.), 89–105. Amsterdam: John Benjamins.

McCarthy, M. 1998. *Spoken Language and Applied Linguistics*. Cambridge: CUP.

McCarthy, M. & Carter, R. 1995. Spoken grammar: What is it and how can we teach it? *ELT Journal*, 49(3): 207–218.

McCarthy, M. & Carter, R. 1997. Written and spoken vocabulary. In *Vocabulary: Description, acquisition and pedagogy*, N. Schmitt & M. McCarthy (eds), 20–39. Cambridge: CUP.

O'Keeffe, A. & Farr, F. 2003. Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly* 37: 389–418.

Quaglio, P. 2004. The language of NBC's *Friends*: A comparison with face-to-face conversation. PhD dissertation, Northern Arizona University.

Quaglio, P. & Biber, D. 2006. The grammar of conversation. In *The Handbook of English Linguistics*, A. McMahon & B. Aarts (eds), 692–723. Oxford: Blackwell.

Stenström, A.-B. 1991. Expletives in the London-Lund corpus. In *English Corpus Linguistics: Studies in honour of Jan Svartvik*, K. Aijmer & B. Altenberg (eds), 239–253. London: Longman.

Waksler, R. 2001. A new *all* in conversation. *American Speech* 76: 128–138.

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

Appendix

Grammatical features selected for analysis*

Grammatical features	Instances selected for analysis
Lexical verbs	say [incl. past tense], get, go, know, think, see, make, come, take, want, give, mean, tell
Copular verbs	become, get, look, feel, seem, go, remain, keep, grow, sound, prove, appear
Expletives	damn, bastard, bitch(y), son of a bitch, shit(ty), fuck (+var), ass, butt, crap(py)
Slang terms	cool, suck, piss(ed)(off), screw(ed)(up), check out, hang out, totally(agreement), what's up?, freak out, I'm out of here, later
Vague language devices	kind of, sort of, or something (like that), or anything (like that), (and) stuff (like that), stuff, shit (expletive used as vague reference)
Inserts	I mean, you know, you see, oh, well, wow, yeah (non-minimal response)
Stance markers	probably, of course, perhaps, maybe, actually
Vocatives (familiarizers)	guys, man, dude, buddy, folks, bro, bud
Non-clausal units	non-clausal questions, wow, exactly, sure, right, fine, good, lovely, okay, uh-huh (+ var)
Intensifiers	so, really, too, totally, damn
Innovations	all + adj/gerund, so + verb, so (not) + NP, so not + adj, totally (agreement), *ish* (time, color, etc), so (at the end of a turn), in vs. for (negative + present perfect + time expression)
Lexical bundles in conversation	I don't know what, I don't know if, do you want to, you know what I, I don't want to
Lexical bundles in Friends	I can't believe you, are you doing here, I want you to, are you talking about, to talk to you, what do you think, thank you so much, what do you mean
Vernacular features	ain't, me and . . ., there's + plural notional subject
Pro verb do + it	do + it, does + it, did + it, have(has) done + it
Aspect, present / past perfect,	simple, progressive, perfect, present perfect, past perfect, past perfect (canonical function), past perfect (counterfactual function)
Verbs controlling that- *&* to-*clauses)*	think that, say that, guess that, want to, try to, like to, seem to
Tense	present tense, past tense
Emphatic do	do, does, did
Modals &semi-modals	necessity/obligation, possibility/permission, prediction, semi-modals
Possibility &permission modals (individually)	can, could, may, might
Personal pronouns	1st person pronouns, 3rd person pronouns
Repeats	2-repeats, 3-repeats, 4-repeats
Greetings & leave-takings	hi, hello, hey, bye + bye-bye, goodbye, (I'll) see you later, (I'll) see you, later, I'm out of here

* Only a few of these features were included in the present article. For the complete analysis of these features, see Quaglio (2004).