Journal of English Linguistics

http://eng.sagepub.com/

Conversational Grammar- Feminine Grammar? A Sociopragmatic Corpus Study

Christoph Rühlemann

Journal of English Linguistics 2010 38: 56 originally published online 4 December 2009
DOI: 10.1177/0075424209347175

The online version of this article can be found at:
http://eng.sagepub.com/content/38/1/56

Published by:
SAGE
http://www.sagepublications.com

Journal of English Linguistics
38(1) 56–87

DOI: 10.1177/0075424209347175
http://eng.sagepub.com
SAGE

Conversational Grammar–Feminine Grammar? A Sociopragmatic Corpus Study

Christoph Rühlemann[1]

Abstract

One area in language and gender research that has so far received only little attention is the extent to which the sexes make use of what recent corpus research has termed "conversational grammar." The author's initial findings have suggested that the majority of features distinctive of conversational grammar may be used predominantly by female speakers. This article reports on a study designed to test the hypothesis that conversational grammar is "feminine grammar" in the sense that women's conversational language is more adapted to the conversational situation than men's. Based on data from the conversational subcorpus of the British National Corpus and following the situational framework for the description of conversational features elaborated in the author's previous research, features distinctive of conversational grammar are grouped into five functional categories and their normed frequencies compared across the sexes. The functional categories distinguish features that can be seen as adaptations to constraints set by the situational factors of (1) Shared Context, (2) Co-Construction, (3) Real-Time Processing, (4) Discourse Management, and (5) Relation Management. The study's results, described in detail in relation to the biological category of speaker sex and cultural notions of gender, suggest that the feminine grammar hypothesis is valid.

Keywords

conversational grammar, conversational situation, sociolinguistics, corpus linguistics, speaker sex, gender, adaptation

Attitudes in corpus linguistics toward context in its many manifestations seem to be undergoing change. While context has for some time been "something of a challenge to the corpus linguist" (Thompson & Hunston 2006:4), it is a key concept in a number

[1]Ludwig-Maximilians-University, Munich, Germany

Corresponding Author:
Christoph Rühlemann, Ludwig-Maximilians-University, 36 Bergerwaldstr., Munich, 81825, Germany
Email: cruehlemann@aol.com

of recent corpus linguistic publications (e.g., Adolphs 2008), suggesting that context is becoming a basic coordinate of corpus linguistic research. This focus on the importance of context in analyzing corpus data is most welcome, but its somewhat belated discovery is surprising, considering that when looking at corpus data we are looking at "used language," or "language in use"—and use inevitably occurs in some context. One large dimension of context that has been given little attention so far in corpus linguistics is the *user*—for, obviously, "use" presupposes a "user." While, in recent decades, sociolinguistic research has accumulated an impressive body of research into how language variation is correlated with social variables, this question has up until now triggered only a small number of studies based on contemporary corpora (whereas historical corpora have been exploited more fully), although "spoken corpora have tremendous potential for the analysis of sociolinguistic variation" (Barbieri 2008:60; also see Rayson, Leech, & Hodges 1997:133).

The list of corpus-based sociolinguistic studies includes the following handful. Rayson, Leech, and Hodges (1997) undertake a comparison of speakers' use of selected vocabulary, examining what words are most frequent in the speech of the sexes, different age groups, and social strata. Andersen (2001) and Stenström, Andersen, and Hasund (2002) analyze in great detail London teenage language recorded in the Bergen Corpus of London Teenage Language (COLT), considering distributional differences regarding age, sex, socioeconomic background, and ethnicity; a more recent study comparing the use of invariant tags in COLT to their use in the Linguistic Innovators Corpus is Torgersen and Gabrielatos (2009). Schmid (2003:186) studies conversational behavior and the use of semantic fields "across the male and female cultures." Barbieri (2007) investigates the effects of speaker age and sex on the use of quotatives in American English, McEnery and Xiao (2004) explore swearing in the British National Corpus (BNC), and Rühlemann (2007) includes a number of sociolinguistic analyses of conversational features. Xiao and Tao (2007) is a sociolinguistic study of amplifiers in British English, and Barbieri (2008) explores age-based linguistic variation in spontaneous American conversation. Although the number of corpus-based studies into social differentiation is relatively small, the clustering of the studies in recent years seems to suggest that corpus linguistics is indeed beginning to exploit the potential of corpora for sociolinguistic studies, and corpus-based sociolinguistics may be about to become another standard branch of corpus linguistic inquiry.

One area, however, in which corpus research has yielded substantial new insights but failed so far to take sociolinguistic concerns into account is the study of conversational grammar. Beside numerous research articles tackling individual features (discussed below in more detail), the grammar of conversation has been described in unprecedented detail in two recently published, corpus-based grammars, the *Longman Grammar of Spoken and Written English* (LGSWE; Biber et al. 1999) and the *Cambridge Grammar of English* (Carter & McCarthy 2006). Particularly revealing are the analyses in the large chapter in the LGSWE titled "The Grammar of Conversation" (Biber et al. 1999:1038-1125). Following the authors of the LGSWE, I use the term *conversational grammar* to mean "grammatical features that are especially

characteristic of conversational language, as compared with other registers" (Biber et al. 1999:1038). That is, conversational grammar is an instance of a "variety grammar" (Rühlemann 2007:15-17): for a feature to be part of conversational grammar, it is required that it be characteristically more frequent in conversation than in other major registers such as academic, fictional, or newspaper writing, to name the three written registers considered in the LGSWE. To illustrate this point, the pronouns *I* and *you* are regarded as part of the grammar of conversation not, of course, because they are nonexistent in, for example, academic writing but because it is in conversation that they are by far most common (cf. Biber et al. 1999:333-334).[1]

Research on conversational grammar has not generally been concerned with, except for a few occasional remarks, the question of whether conversational grammar might covary not only with register—a category according to "use"—but also with social categories—which are categories according to "user." The present study aims to examine the extent to which the use of conversational grammar is differentiated according to one such category, speaker sex. The overriding research question this article addresses is the following: Is speaker sex correlated with the use of conversational grammar?

In addressing this question we need to distinguish between speaker sex and gender, a distinction reflected in the word pairings female/male and feminine/masculine. While sex relates to being female and male, which are biological categories, gender is related to the notions of "femininity" and "masculinity," which are socially constructed. In West and Zimmermann's (1987:127) words, sex is "a determination made through the application of socially agreed upon biological criteria for classifying persons as either females or males." Gender, in contrast, is "the activity of managing situated conduct in light of *normative* conceptions of attitudes and activities appropriate for one's sex category" (West & Zimmermann 1987:127, emphasis added). Corpora such as the BNC, containing very large amounts of naturally occurring text, are maybe at their best when used to reveal patterns that occur en masse because these may be what is typical and, hence, the (quantitative) "norm" in a language community. The aim of this corpus study is to examine highly frequent features distinctive of conversation in terms of their distribution across the sexes, thus potentially revealing normative attitudes toward the use of conversational grammar. For example, if it could be shown that conversational grammar tends to be used much more frequently by men, this might reveal the normative conception that using conversational grammar is part of masculinity—something appropriate for men to do but less so for women. And the reverse: if it turned out that female speakers use conversational grammar much more than male speakers, this would suggest the normative conception that using conversational grammar is feminine—something women appropriately do but not men.[2] The raw material this study is concerned with is speaker sex; the observations made, however, may have important implications for how speakers do gender.I single out speaker sex as the target variable of this study on two grounds. First, in the past thirty years or so, an immense body of sociolinguistic research has established the notion of sex differentiation "as a recurrent robust finding" (Holmes 1997:197). Therefore, it seems

highly likely that sex differentiation will also affect the use of conversational grammar. Second, Rühlemann (2007) found initial evidence suggesting that conversational grammar might be subject to variation across men and women in that all features whose distribution across the sexes was examined were predominantly used by women. Building on this work, I hypothesize that conversational grammar is "feminine grammar," with "feminine" understood in the above-cited sense of what is conceived of as normative and thus appropriate for women. The aim of this article is to test the feminine grammar hypothesis, according to which women's conversational language is more adapted to the conversational situation than men's.

To fully appreciate this hypothesis and its implications, it is necessary to review two notions the hypothesis involves: conversational situation and adaptation. These two notions are briefly explained in the following section.

Situation and Adaptation

Conversation, seen as a register in the Hallidayan sense, is intimately linked to the notion of situation or, more specifically, "situation type" (Halliday 1978:28), which refers to not the apparently infinite number of different possible situations but the general type of situation that gives rise to conversation. What is this conversational situation like? Various scholars have recently attempted to address this question (e.g., Biber et al. 1999; Leech 2000). The account of the conversational situation I present largely builds on Rühlemann (2007), probably the most comprehensive attempt to capture the factors determining the conversational situation type. Rühlemann outlines five factors: Shared Context, Co-Construction, Real-Time Processing, Discourse Management, and Relation Management.

Shared Context refers to the wealth of perceptual, nonverbal, and social context that conversationalists (who are typically socially familiars) share with each other. Co-Construction reflects the fact that conversational text is inevitably constructed jointly, through sequential organization (Sacks, Schegloff, & Jefferson 1974) and role rotation (Goffman 1981). Real-Time Processing captures the fact that conversationalists are fully exposed to pressures of planning, processing, and executing speech in real time. Discourse Management describes the fact that, because of Co-Construction and Real-Time Processing, conversationalists need to cope with an unparalleled dynamism both in interactional terms—who talks when and for how long is not predetermined—and in cognitive terms—what is talked about is not predetermined either. Finally, Relation Management refers to the type of goal orientation, which is in conversation decidedly interpersonal, with the overall goal being to "establish bonds of communion" (Malinowski 1923).

The notion of adaptation comes into play where conversational grammar, as the set of forms distinctive of the register of conversation, is seen in the context of a "textual pragmatics" (Leech 1983), that is, in a perspective in which grammar is regarded as being "under the functional influence of pragmatics" (Leech 1983:64). Looked at from this perspective, grammatical choices in actual discourse are less "rule

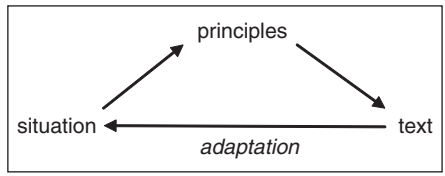

Figure 1. A model of adaptation in a "textual pragmatics"

Table 1. Situational Factors and Corresponding Principles

Situational Factor	Corresponding Principle
Shared Context	Interactivity
Co-Construction	Context reliance
Real-Time Processing	Economy
Discourse Management	Processibility
Relation Management	Expressivity politeness

governed"—that is, less motivated by speakers' compliance with formal grammatical rules—than "principle controlled" (Leech 1983:5)—that is, motivated by speakers' communicative goals and constraints—whereby certain pragmatic principles help adapt the form of text to constraints set by the situation, as shown in Figure 1.

In the remainder of this section I provide a rough sketch of the principles controlling conversational text vis-à-vis the factors determining the conversational situation. They are listed in Table 1.

As noted above, Shared Context refers to the variety and richness of contextual dimensions that typically characterize the conversational situation, including (1) the perceptual, or spatial, context the physical setting provides, (2) the social context the speakers' being familiar with each other provides, and (3) the nonverbal "micro-context" (Arndt & Janney 1987) the vocal and kinesic channels of communication provide. The principle controlling language use vis-à-vis this wealth of Shared Context can be described as Context Reliance, that is, as a tendency to "outsource" meaning to the unsaid. Context Reliance favors linguistic phenomena such as deixis in its various forms (which is commonly exophoric or text external),[3] anaphora (which is endophoric or text internal), situational ellipsis, vague language, multimodality—the fact that "cognitive, conative and emotive information can be signaled in any of the three basic modes [verbal, vocal and kinesic]" (Arndt & Janney 1987:395)—and crossmodality—"the fact that signals in one mode can to a certain extent be substituted for or replaced in other modes" (Arndt & Janney 1987:395).

The principle related to the situational factor of Co-Construction is Interactivity. Since the conversational situation generally prevents holding the floor for extended periods (the most notable exception being narrative) and, instead, demands that at

least two partners jointly weave the text, the Interactivity Principle ensures that turn taking and, as a concomitant, speaker change occur sufficiently and that conversational flow is maintained in a reciprocal manner. The processes favored by this principle include co-constructing utterances (e.g., Lerner 1991; Helasvuo 2004), backchannelling (e.g., Holmes & Stubbe 1997; Wong & Peters 2007), the use of question tags and response elicitors (e.g., Biber et al. 1999; Tottie & Hoffmann 2006), and the use of certain pragmatic markers such as *like*, whose ultimate function is to mark discourse as "analogous" (Adolphs & Carter 2003) or "loose" (Andersen 1998, 2001) and which "leaves the conversational partner room to complete or edit internally and to respond and interact, thus inviting the collaboration of the recipient in the negotiation of meaning" (Rühlemann 2007:147).

The principle that helps conversationalists adapt to the scarcity of planning and processing time (Real-Time Processing) is the Economy Principle. Leech (1983:67) describes it as follows:

> The Economy Principle ("Be quick and easy") can be regarded as a valuable precept not only for *h* [the hearer] but also *s* [the speaker]. If one can shorten the text while keeping the message unimpaired, this reduces the amount of time and effort involved both in encoding and in decoding.

The processes that the Economy Principle favors include forms of reduction both on the phonological and the syntactic level. Phonologically reduced forms include verbal and negative contractions, conversational contractions such as *gonna* and *gotta*, and forms of situational ellipsis (cf. Quirk et al. 1985:896). Syntactic reduction is brought about by deletion (e.g., zero relativizer both in object and subject position) and substitution (e.g., pronominalization and use of other proforms such as predicative *so*).

Discourse Management, understood as the need conversationalists feel to cope with the interactional and cognitive dynamism the conversational situation creates, is responded to by the Processibility Principle. This principle "recommends that the text should be presented in a manner which makes it easy for the hearer to decode in time" (Leech 1983:64). Easing processing for the recipient is achieved in conversation by means of, for example, what Biber et al. (1999:1068) term "syntactic analysis"—that is, by headers and tails (e.g., Aijmer 1989; Carter & McCarthy 2006); by means of the use of discourse markers, which act as "discourse glue" (Fraser 1990) providing discourse coherence (Schiffrin 1987); and by means of the "maxim of end-weight" (Leech 1983:65) which demands that complex constituents be right branched rather than left branched (e.g., the preference in speech for extraposition with *it*; see Kaltenböck 2004).

The principles correlated with the fifth situational factor, Relation Management, are more diffuse and harder to define because the realization of the interpersonal goal orientation through language is "a domain of low certainty and high complexity" (Arndt & Janney 1991:536). A broad outline is therefore all that is possible in this article. For the present, at least two principles seem to be involved in mediating

between conversational language use and the interpersonal goal orientation: the Expressivity Principle and the Politeness Principle. Speaking of certain types of repetition, Leech (1983:69) notes that "the emphasis of repetition has some rhetorical value such as surprising, impressing, or rousing the interest of the addressee." This description applies to many other phenomena of language use such as hyperbole (cf. McCarthy & Carter 2004), certain types of vocatives (cf. Leech 1999; McCarthy & O'Keeffe 2003; Rühlemann 2007:184-189), and both epistemic and affective stance (e.g., Biber & Finegan 1988, 1989; Biber et al. 1999; Hunston & Thompson 2000), to name only a few. In keeping conversational interaction "on course" toward its primary goal of creating rapport, another important principle is the Politeness Principle. In Leech's interpretation, this principle has a higher regulative role than the cooperative principle postulated by Grice (1975) in that the Politeness Principle essentially serves "to maintain the social equilibrium and the friendly relations which enable us to assume that our interlocutors are being cooperative in the first place" (Leech 1983:82). The processes in conversation favored by the Politeness Principle include, for example, mitigation, crossmodality, and indirectness.

While this is admittedly a preliminary description of the interrelation of situational factors and pragmatic principles, the key point is that conversational grammar needs to be seen in the context of its situation to be appreciated as functional—functional in the sense that conversational language is adapted to constraints set by the conversational situation. In the following I investigate whether indeed the feminine grammar hypothesis can be confirmed, according to which more such adaptation can be observed in women's conversational language use than in men's.

Data and Method

The data for this study come from the demographically sampled subcorpus of the 100-million-word BNC (World Edition). This subcorpus consists of 4.2 million words. It is sometimes also referred to as the "conversational" subcorpus because, as is widely agreed, its 153 "texts"—that is, the transcripts made from various, extended spoken interactions the recruits recorded over a period of several days—"consist of casual conversations" (Aston & Burnard 1998:28; also see Rayson, Leech, & Hodges 1997; Biber et al. 1999:133). To ensure representativeness the creators chose to employ "demographic sampling of the kind which will be familiar from its use in public opinion research, that is, selecting informants on the basis of their age, sex, region, social class, and so on" (McEnery & Wilson 1996:65). A total of 148 recruits were selected in such a way that roughly equal numbers were achieved in terms of recruits' sex, age group, region, and social stratum. Of significance in the present connection is the "small built-in bias" with regard to recruit sex: as recruits, 75 women but only 73 men were selected (Rayson, Leech, & Hodges 1997:135). More importantly, this bias "spilled over" to the total numbers of male and female speakers respectively—that is, recruits *and* their interlocutors; probably due to a common tendency to build same-sex networks,[4] the number of female speakers rose to a total of 561 compared to 536 male

$$\frac{NF\ f - NF\ m}{NF\ f + NF\ m}$$

Figure 2. Difference coefficient formula
Source: Adapted from Leech and Fallon (1992).

speakers. Moreover, women in the conversational subcorpus were not only more numerous but also more talkative: Rayson, Leech, and Hodges (1997:136) report that the number of words per turn was 9.53 for men while it was 10.33 for women. The greater female verbosity is also reflected in the numbers of utterances: men produced 181,255 utterances, and women produced 255,925 utterances.

To neutralize this female overrepresentation it was paramount in this study not to rely on absolute frequencies alone but also to use normalized frequencies per 1,000 utterances. Raw frequencies were nonetheless useful in determining which features to investigate in that only features that have overall (i.e., sex independent) raw frequencies greater than 1,000 occurrences were included. This cutoff point was set to ensure a high rate of statistical significance. A total of 110 features distinctive of conversational grammar were thus investigated. Each feature was assigned to one of the five situational categories detailed above. Sex-sensitive queries were conducted, and raw frequencies for men and women were determined. Chi-square tests were carried out for each feature to establish whether the differences in occurrence between men and women were due to chance or, rather, a reflection of differences in the target population (which, in this case, happens to be the population of Britain). Building on the raw frequencies of occurrence, normed frequencies per 1,000 utterances by men and women were calculated. Since many of the investigated features have very high frequencies and others have comparatively lower frequencies, it was felt necessary to adapt Leech and Fallon's (1992) Difference Coefficient Formula, whereby the difference between the normed frequencies per 1,000 utterances by women (NFf) and men (NFm), respectively, is divided by their sum. The formula is shown in Figure 2.

The Difference Coefficient Formula is extremely useful in that it effectively compresses differences of occurrence, which may vary immensely between the features, into values ranging between 0 and +/–1, thus allowing direct comparisons across the features. When a positive value is obtained, the feature in question is more frequent for women; when a negative value is obtained, the feature is more frequent for men. The (hypothetical) boundary coefficients +1.00 and –1.00 would indicate that a feature is used only by women or only by men, respectively.

Many of the features investigated are highly multifunctional. Care was therefore taken to formulate the corpus queries in such a way that only the functions under investigation were included. To illustrate, the phrase *you know* can fulfill at least three distinct functions. The verb *know* can be used transitively and, thus, in its full lexical sense, as in *How do you know it isn't on?* This usage is not specific to conversation but can be found in many registers.[5] It is therefore not investigated in this study. By contrast,

when *you know* is spoken with a rising intonation and in utterance-final position, as in *I like to check that it all works you know?,* it can serve as a type of invariant tag question. Since tag questions are distinctive of conversation, this usage would qualify for inclusion in this study. However, a query for *you know?* (with question mark) returns only 850 occurrences, thus falling short of the 1,000 occurrence threshold. Finally, the most frequent function of *you know* in conversation is as a discourse marker, whereby a speaker signposts his or her discourse for the listener as being "readable" against the backdrop of shared background knowledge. A clear distinction between the three functions is not always possible, as in, for example, *But you know I've been terribly busy lately!* where only inspection of concordance lines could disambiguate whether *you know* performs a transitive verb function or a discourse marker function. To exclude instances of the two non-discourse-marking functions and to circumvent ambiguous cases, the query was for *you know* bracketed by commas because in the BNC transcription commas represent slight terminal rises followed by a short pause (Crowdy 1994:26). An example is *Let's just, you know, open the door*. Other methods used to make the queries as "precise" as possible included searching for features via part-of-speech tags (e.g., the generalized use of *me* as a possessive determiner is tagged in the BNC as DPS), restricting queries to features in certain positions within utterances (e.g., the notoriously multifunctional item *well* is most likely to perform a discourse marker function when used in utterance-initial position), and performing queries for items as stand-alone features (laughter as stand-alone item is most likely to perform a backchannel function). Obviously, this methodology is far from perfect because, to return to *you know*, many instances of discourse marker *you know* are spoken without intonational brackets and have therefore been transcribed without commas; these other instances are hence "lost" for this study. Nonetheless, given that the study is functional in the sense that a major focus is placed on investigating conversational features as adapted, and thus *functionally related*, to the conversational situation, a trade-off between recall and precision that favors the latter was seen as indispensable.

Results

Shared Context

Thirty-two features were assigned to the situational factor Shared Context. They are divided in four broad categories: (1) deixis, including the three traditional deictic subcategories person, time, and place deixis, (2) anaphora, (3) vague language, and (4) paralinguistic behavior. Deictics such as *I*, *you*, or *today* and *here* were included because personal, temporal, and locational deictics have been found to be most frequent in conversation (Rühlemann 2007) and because, for example, in Kilgarriff's (1998) frequency list, *I* and *you* are the two most frequent words in the conversational subcorpus. The role of anaphora is similarly prominent in conversation. Conversation, contrary to much writing, which relies on spelling out reference, exhibits a preference

for taking reference for granted, a preference that is manifested in the much more frequent use of anaphoric third-person pronouns in conversation (again, see Kilgarriff's frequency list). Vague language, such as the "place holder" *thing(s)* and the "set marker" *or something* (Stenström, Andersen, & Hasund 2002), as in *Didn't Rob get a job there or something?*, activates the audience's knowledge of the world and achieves communication through not specific but rather vague reference. That is, what is ultimately communicated is much more than what has been made explicit: listeners fill in the gaps with knowledge from the large, varied context they share with the speakers (cf. O'Keeffe 2004). Paralinguistic phenomena, finally, such as laughter as a voice quality—that is, laughter not between speech but within speech uttered in a laughing tone (Aston & Burnard 1998:171)—is a prime example of how nonlinguistic communicative means interact with linguistic ones, providing a micro context (Arndt & Janney 1987) in the light of which speech is interpreted.[6]

Table 2 lists all Shared Context features investigated in this study. Like Tables 3 to 6 below, it details the normed frequencies for men and women as well as the difference coefficient values obtained from the Difference Coefficient Formula (see Figure 2). It also contains the chi-square values obtained (χ^2); values in bold indicate differences in frequency of occurrence between men and women that are significant at the $p < .01$ level. Note that for one degree of freedom ($df = 1$) the critical value for $p < .01$ is 6.63. That is, for example, the differences in the frequencies between men and women obtained for the pronoun *I* are significant at $p < .01$ because the chi-square value (1921.43) obtained is far greater than the critical value 6.63; the differences for the pronoun *us*, by contrast, which yielded a chi-square value of 0.7, are not significant at that level.

In toto, twenty-nine out of thirty-two Shared Context features (91 percent) are more frequent for women. The differences observed are significant at the $p < .01$ level for twenty-seven features, only one of which is more frequent for men.

As regards deictics and anaphora, Table 2 shows that the person and time deictics as well as anaphoric pronouns are consistently more frequent for women. By contrast, the place deictic *here* is more frequent for men. This fact is potentially interesting in that it is reminiscent of Rayson, Leech, and Hodges's (1997:140) observation that "men show a greater predilection for place names than women." Admittedly, *here* is not a place name, and the differences regarding *here* are not significant at the $p < .01$ level. We must therefore be wary of drawing far-reaching conclusions. It is nonetheless tempting to hypothesize that the use of deixis may depend on sex: women, using more person and time deixis, might orient more strongly to the personal and the temporal dimensions of the context of situation, whereas men, who seem to be using more place deixis, might take their bearings primarily from the spatio-locational dimension of the speech situation. Although merely tentative, this hypothesis merits further exploration.

Another striking observation is that the coefficient values for all feminine pronouns—*she* (0.39), *her* (tagged PNP; 0.37), and *her* (tagged DPS; 0.40)—are much more frequent for women than for men (cf. Rayson, Leech, & Hodges 1997:138). This might be taken as revealing a tendency for women not only to build same-sex social

Table 2. Shared Context Features

Feature		NFm	NFf	Coeff.	χ^2
Deixis					
Person deixis	I (PNP)	298.26	361.79	.10	**1921.43**
	you (PNP)	250.13	283.65	.06	**605.26**
	we (PNP)	63.33	68.83	.04	**51.60**
	me (PNP)	26.13	31.81	.10	**119.57**
	your (DPS)	25.51	31.81	.11	**148.31**
	my (DPS)	20.12	25.72	.12	**145.41**
	us (PNP)	6.69	6.76	.01	0.07
Time deixis	now	22.70	24.46	.04	**14.13**
	today	4.08	5.47	.15	**41.78**
	tomorrow	3.01	3.96	.14	**26.70**
	yesterday	1.74	2.81	.24	**51.07**
	tonight	1.99	2.51	.12	**12.30**
	this morning	1.77	2.64	.20	**34.95**
Place deixis	here	17.40	17.21	–.01	0.21
Anaphora					
Pronouns	it (PNP)	247.70	268.11	.04	**229.91**
	he (PNP)	86.92	104.25	.09	**363.53**
	they (PNP)	82.90	92.81	.06	**128.83**
	she (PNP)	38.77	89.09	.39	**4227.59**
	them (PNP)	29.30	34.27	.08	**83.95**
	him (PNP)	14.83	20.31	.16	**179.63**
	his (DPS)	12.05	13.79	.07	**24.79**
	her (PNP)	7.73	16.84	.37	**682.67**
	her (DPS)	5.03	11.70	.40	**532.51**
	their (DPS)	5.04	5.50	.04	4.17
Vague language					
Set marker	or something	3.33	3.54	.03	1.30
Vague words	some	14.57	16.70	.07	**30.80**
	kind/sort/type of	8.69	9.82	.06	**14.51**
	a bit	7.68	9.77	.12	**52.29**
	about (ADV)	7.25	6.98	–.02	1.06
	whatever	2.76	2.15	–.12	**16.21**
Place holder	thing/s	16.13	17.67	.05	**14.94**
Paralinguistics					
Laughter	laughing (voice qual.)	9.30	15.29	.24	300.70

Note: NFm = normed frequency/1,000 utterances by men; NFf = normed frequency/1,000 utterances by women; coeff. = difference coefficient value; χ^2 = chi-square value ($p < .01$, $df = 1$, critical value = 6.63).

networks, as noted above, but also to address same-sex related topics. However, women also use the male pronouns *he*, *him*, and *his* more frequently than men; yet the difference coefficients obtained for male pronouns are invariably much lower than the ones obtained for female pronouns. This overrepresentation of both female and male pronouns for women ties in well with Rayson, Leech, and Hodges's (1997:140)

observation that "women use personal names considerably more than men do." Women's preference for female and male pronouns and for personal names perhaps bear out Rayson, Leech, and Hodges's (1997:139) hypothesis that "female speech is more interactive and concerned with establishing and maintaining relationships."

The picture is less clear as regards vague language. Men use the vague items *about* (ADV; though insignificantly) and particularly *whatever* more often, while women show a preference for the set marker *or something* (though insignificantly) as well as for *some*, *kind/sort/type of*, and particularly the place holder *thing/s*. Striking differences are found with regard to laughter: women use much more laughing voice quality than men (coefficient = .24). Note that in Rühlemann (2007) and Günther (2003), women were also observed to use much more laughter as a vocal event (between-speech laughter).[7]

Co-Construction

The features that can be seen as adaptations to constraints set by the factor Co-Construction fall into three broad categories: turn taking, backchannels, and negation. Contrary to Sacks, Schegloff, and Jefferson (1974) who postulate the no gap/no overlap rule for the sequential organization of conversation, some studies have shown conversation to exhibit a large amount of overlap (e.g., Coates 2006; Rühlemann 2007:38-39). Despite the indubitable functional complexity that characterizes questions as well as question tags as a question subtype (Quirk et al. 1985:806; for an overview of the manifold subfunctions of question tags, see Tottie & Hoffmann 2006), it is assumed in this study that the underlying function of both questions and question tags is to oil the wheels of turn taking and hence "to foster linguistic interaction" (Schmid 2003:196). Among the backchannels included in this study is laughter coded in the BNC as a vocal event (between-speech laughter), which, particularly when used as a stand-alone item, serves typical backchannel functions (cf. Rühlemann 2007:80-86). That backchanneling behavior has a quintessential role in the co-construction of conversational text has very recently been demonstrated by Wong (2008), who found that turns that received backchannel feedback were, on average, five times longer than turns without such feedback. As far as negation and its co-constructive role is concerned, a number of studies have shown that negation is particularly frequent in conversation compared to other registers (e.g., Tottie 1991; Biber et al. 1999:159; Leech 2000:696), a phenomenon that, as Leech (2000:696) comments, "reflects the speakers' tendency to interact through contrastive perspectives" (for the association of negation with question tags and mental verbs, see Tottie 1991; Rühlemann 2007).

Table 3 displays all forms investigated as well as the coefficients and chi-square values obtained. Note that, for the backchannels, the frequencies of occurrence shown in the table refer to their occurrences as stand-alone items (i.e., utterances consisting of *yeah*, *laugh*, or *oh* only) and, in the cases of *mm* and *mhm*, as forms in utterance-initial position only.[8]

As shown in Table 3, fourteen features were investigated in the Co-Construction category: eleven are found to be more frequent for women compared to three that

Table 3. Co-Construction Features

Feature		NFm	NFf	Coeff.	χ^2
Turn taking					
Overlap	<ptr> (u-in.)	72.05	72.86	.01	1.02
Questions	?	201.45	218.67	.04	**188.61**
Tags					
Question tags	is it?	4.79	5.15	.04	2.69
	isn't it?	4.37	5.02	.07	**9.36**
	innit?	3.04	3.55	.08	**8.14**
Invariant tag	eh?	2.70	2.02	–.14	**21.28**
Backchannels	mm (u-in.)	29.95	38.07	.12	**208.68**
	yeah (stand-alone)	12.43	11.02	–.06	**18.28**
	laugh (stand-alone)	8.09	10.26	.12	**53.66**
	mhm (u-in.)	3.18	3.09	–.01	0.25
	oh (stand-alone)	2.13	2.54	.09	**7.38**
	BE_? (u-in.)	2.00	2.26	.06	3.21
	DO_? (u-in.)	1.73	2.12	.10	**8.07**
Negation	n't	133.55	173.41	.13	**1274.72**

Note: NFm = normed frequency/1,000 utterances by men; NFf = normed frequency/1,000 utterances by women; coeff. = difference coefficient value; χ^2 = chi-square value ($p < .01$, $df = 1$, critical value = 6.63).

are more frequent for men. Looking at the statistically significant differences only, there are ten features total, eight of which are more frequent for women, accounting for 80 percent.

As to turn-taking features, overlap (indicated in the markup used in the BNC by <ptr>-elements), questions (indicated by question marks), and all three types of question tags investigated (*is it?*, *isn't it?*, and *innit?)* are more frequent for women. Conversely, the invariant tag *eh?* was clearly more frequent for men.

The picture is rather mixed with regard to backchannels: while utterance-initial *mm*, stand-alone laughter (as a vocal event), stand-alone *oh*, and "follow-up tag questions" (cf. Carter & McCarthy 2006) in utterance-initial position (e.g., *Are they?* or *Does he?)* are more frequent for women, stand-alone *yeah* and utterance-initial *mhm* are more frequent for men, but by only slight margins. This finding runs counter to some previous research making sweeping claims that "it is men rather than women that fail to respond minimally" (Coates 1993:112) and, instead, supports alternative research suggesting that backchannel behavior is gendered in a much more differentiated and subtle way (see, e.g., Reid 1995).

Real-Time Processing

The group of features that can be seen as adapted to the fundamental scarcity of planning and production time that constrains talk in conversation and that were investigated for the present study includes four broad categories. Features such as silent and filled pauses realized through *er* and *erm*, often referred to as "hesitators," belong to the

group of what Allwood, Nivre, and Ahlsén (1990) term "speech management phenomena" (also referred to as "disfluency" features). Another feature included under speech management is repetition of *I*, as in *But I cert I I I I I ju I it it just sounds*. . . . As Stenström and Svartvik (1994) show, *I*-repetition concerns what they call "grammatical subject territory" and subjects occur utterance initially where online planning is building up. *I*-repetition is thus a direct consequence of real-time constraints (cf. Biber et al. 1999:334). Part of the investigation were also two types of contraction: verbal contractions such as *'s* (for *is*, tagged VBZ) and *'ll* (for *will*) and conversational contractions such as *gonna*, *gotta*, and so on. Furthermore, the Real-Time Processing category includes morphosyntactic variants such as possessive *me* (as in *She says me dad's Turkish*),[9] the quotative forms *I says/goes* (as in *Steve says to me, is he in? I says, no. He says, he's not in? I says, no.*) and use of *I/he/she/it were* (as in *That's what I were doing, twice*).[10] The common denominator is that, unlike morphological variants such as *yous* (cf. Biber et al. 1999:1123) which are forms outside the Standard English canon, these morphosyntactic variants involve Standard English forms such as *me*, *goes*, and *were* but generalize them to nonstandard grammatical functions; for example, the personal pronoun *me* is generalized to the function of the possessive determiner *my*. The forms are hence subsumed under Generalization. It has been shown that particularly the generalized quotative forms *I says* and *I goes* can be seen as adaptations to real-time processing constraints in that these forms are both morphologically and phonologically aligned to the forms they most frequently alternate with, namely *he/she says* (Rühlemann 2007:169-180) and *he/she goes* (Rühlemann 2008; on the effect of frequency on phonological reduction, also see Bybee & Scheibman 1999).[11]

In sum, in the Real-Time Processing category, fourteen out of nineteen features (74 percent) are more frequent for women. The differences in frequency between men and women are found significant for sixteen features, of which twelve are predominant in women's conversation, accounting for 75 percent.

As can be seen from Table 4, verbal contractions and generalized forms are consistently more frequent for women, with generalized forms reaching very high coefficients (ranging between 0.27 and 0.39). The results both for speech management features and conversational contractions are mixed: silent pauses are equally distributed between the sexes; the filled pauses *er* and *erm* show contrary behaviors in that *er* is more frequent for men while *erm* is more frequent for women. Verbal contractions seem to be more frequent in men's conversations, with *gonna* just slightly more frequent for women but *gotta*, *wanna*, and *dunno* clearly more frequent for men.

The findings on filled pauses are in stark contrast to previous research by Schmid (2003:193) who notes higher frequencies of both *er* and *erm* for men than for women, a difference that may be due to the fact that Schmid's database is the whole of the spoken subcorpus of the BNC which includes not only the conversational subcorpus but also the "public speech" recorded in the context-governed subcorpus. The findings suggest that planning constraints, as evidenced by silent and filled pauses as well as *I*-repetition, make themselves felt roughly equally for men and women.

Table 4. Real-Time Processing Features

Feature		NFm	NFf	Coeff.	χ^2
Speech management	pause	291.56	291.31	.00	0.03
	er	51.91	36.46	–.17	**617.15**
	erm	28.41	35.43	.11	**166.05**
	I (repetition of)	2.54	1.91	–.14	**19.15**
Verbal contractions					
BE	's (VBZ)	175.63	185.55	.03	**70.26**
	're (VBB)	37.04	41.35	.05	**51.78**
	'm (VBB)	26.46	28.77	.04	**20.86**
HAVE	've (VHI)	41.47	49.06	.08	**139.40**
	's (VHZ)	18.41	24.99	.15	**210.84**
	'd (VHD)	4.56	7.33	.23	**132.02**
will	'll (VMO)	36.49	41.04	.06	**58.29**
would	'd (VMO)	10.20	12.61	.11	**53.50**
Conversational contractions					
	gonna	15.93	16.17	.01	0.37
	gotta	5.67	5.60	–.01	0.08
	wanna	4.59	3.53	–.13	**29.92**
	dunno	3.32	2.76	–.09	**10.98**
Generalization					
	me (DPS)	2.11	3.86	.29	**103.35**
	I says/goes	1.32	3.02	.39	**132.30**
	I/he/she/it+were	2.12	3.71	.27	**87.78**

Note: NFm = normed frequency/1,000 utterances by men; NFf = normed frequency/1,000 utterances by women; coeff. = difference coefficient value; χ^2 = chi-square value ($p < .01$, $df = 1$, critical value = 6.63).

Discourse Management

Features belonging to three major categories were investigated as Discourse Management features: discourse markers, such as *cos* and *like* (tagged ADV; as in *Cos, well the floor can be a disco anyway, like, you know.*), whose principle function is the establishment of discourse coherence (Schiffrin 1987); quotatives such as *said* and TELL, which act as turn markers in conversational discourse presentation (cf. Rühlemann 2007); and what Biber et al. (1999:1118) term "utterance openers," that is, small words such as *oh* and *well* (as in *he said well I need it*.) that function as auditory quotation marks signaling that speakers are embarking on direct mode quotation.

As is shown in Table 5, the fourteen Discourse Management features investigated are invariably more frequent for women. The differences between women and men are insignificant for two features (*you see* and *, you know,*); so twelve features are significantly more frequent for women than for men.

As far as discourse markers are concerned, the differences observed are small (and insignificant) for *you see* and , *you know,* but large and significant for *cos* and *because*, two seeming variants that in fact fulfill different functions in discourse: "*Because* is

Table 5. Discourse Management Features

Feature		NFm	NFf	Coeff.	χ^2
Discourse markers	well (u-in.)	32.94	37.83	.07	**73.38**
	cos	18.27	26.82	.19	**340.93**
	I mean	17.76	21.54	.10	**77.27**
	because	10.09	15.24	.20	**217.29**
	like (ADV)	9.14	11.27	.10	**46.64**
	you see	6.87	6.93	.00	0.05
	, you know,	4.32	4.60	.03	1.80
	so (CONJ)	3.03	3.56	.08	**8.80**
Quotatives	said (VVD)	25.62	46.39	.29	**1259.05**
	TELL	11.50	14.88	.13	**90.92**
	says (VVZ)	7.43	10.40	.17	**102.76**
	ASK	4.27	5.49	.13	**31.63**
Utterance openers	SAY+well	1.96	3.20	.24	**60.51**
	SAY+oh	1.79	2.77	.21	**42.76**

Note: NFm = normed frequency/1,000 utterances by men; NFf = normed frequency/1,000 utterances by women; coeff. = difference coefficient value; χ^2 = chi-square value ($p < .01$, $df = 1$, critical value = 6.63).

typically used as a subordinating conjunction, while *cos* is the typical discourse marker, often simply functioning as a continuation signal" (Stenström 1998:143; cf. Schleppegrell 1991). Very clear differences are found in quotative usage: not only the lemmas TELL and ASK but also the forms *says* and especially *said* are much more frequent for women than for men. Given this overrepresentation of quotatives it is not surprising that utterance openers, which operate in close association with quotatives, are more frequent for women as well.

The predominance of quotatives and utterance openers for women lends additional evidence to the growing body of research suggesting that women use more discourse presentation. Analyzing a small corpus of narratives told by white middle-class Midwestern Americans, Johnstone (1993:73) found that "when women report speech they do so at greater length and more often in the story [than men]." This observation is supported by research by Ferrara and Bell (1995:274), in whose (larger) corpus of narratives "61% of the males included dialogue in the personal narratives whereas a larger percentage of the females, 74%, included direct speech in narratives." The greater use of discourse presentation by women has also been noticed in corpus linguistic research. In Rayson, Leech, and Hodges (1997:137), a BNC-based study, *said* ranked third in the list of words most "characteristic" of female speech, while in the corresponding male-speech list no quotative verb was included. Stenström, Andersen, and Hasund (2002:126), working on COLT, found that "the girls use more quotative verbs than the boys," while Barbieri (2007), working on data from the American English Conversation component of the *Longman Spoken and Written English* corpus (Biber et al. 1999), found strikingly higher rates of quotative usage among young females up to age twenty-six and females older than forty (only men aged twenty-seven to forty scored higher than women of that age group).

Why is it, then, that women use more discourse presentation? One possible explanation is Tannen's (1990) observation that women show a greater concern not only with the details of who said what but *generally* with any details of past encounters because this concern "shows caring and creates involvement" (Tannen 1990:115). Although this sounds intuitively plausible, the female predilection for discourse presentation is undoubtedly a more complex phenomenon that merits further investigation.

The higher frequencies for women of *says* and *said*, which predominantly serve to introduce direct mode discourse presentation (cf. Rühlemann 2007:127-139), as well as the higher female rates for SAY immediately followed by utterance openers, which invariably signal that speakers "are embarking on direct speech quotation" (Biber et al. 1999:1118), suggest that women use not only more discourse presentation overall but specifically more discourse presentation in direct mode. This observation is significant in terms of deixis: in indirect mode, all deictic features are appropriate to the speaker in the posterior, discourse presenting, situation; in direct mode, all deictic features are appropriate to the speaker in the anterior, presented, situation. Like actors, presenters using direct mode lend their voice to nonpresent speakers, thus slipping into their roles and temporarily assuming their identities. Such role switches can be numerous and succeed each other rapidly, for example, in extended narrative. Moreover, only direct mode can capture "the emotive affective aspects of speech. Insofar as these are expressed not in the content, but in the form of the message, they are not preserved in indirect reporting" (Romaine & Lange 1991:240). That is, it is only in direct mode presentation that the expressive potential of the human voice can be exploited. Thus, direct mode turns discourse presentation into drama. It seems reasonable to assume that this dramatic reenactment not only reflects the presenter's involvement in reliving (portions of) the anterior situation but also serves to increase the vividness of the presentation for the listener. If women use this strategy more frequently than men, this seems to indicate a more flexible attitude toward identity in discourse, which is conceived of not as static and simple but dynamic and multiple and a greater concern for the dramatization of discourse presentation and its engaging effects on the listener.

Relation Management

Given that the relational goal orientation is foundational in conversation, it is unsurprising that the group of features subsumed under Relation Management is large. The thirty-one features studied are assigned to three broad categories: politeness (cf. Leech 1983; Brown & Levinson 1987), epistemic stance, and affective stance (cf. Biber & Finegan 1988, 1989).

Politeness features fall into two subcategories: hedges and mitigators. Hedges include features such as *I (don't) think* or *perhaps* that are used to help the speaker avoid sounding (too) assertive, which might be taken as threatening the addressee's negative face (his or her freedom from imposition).[12] Mitigators, on the other hand, include features such as the phrase *I don't know* which is often used not to declare

insufficient knowledge but as a preface to an act of disagreement or negative assessment. Its pragmatic function is to reduce "the speaker's commitment to the truth of the proposition expressed, hence softening the face-threatening effect of the disagreement or negative assessment" (Diani 2004:169; also see Bybee & Scheibman 1999). Mitigators also include adverbial *though* (as in *I'll have a cup of tea though Alb.*), which "provides a means of disagreeing in a less direct way than *but* or *however*" (Conrad 2004:72); the politeness particle *please*, which occurs mainly in requests (cf. Wichmann 2004:1521) and serves to minimize the imposition implied in requests; and the pragmatic particle *let's*, which can be used as a "crypto-directive (camouflaging an authoritative speech act as a collaborative one)" (Biber et al. 1999:1117; also see Tannen 1990), as in *Richard, let's get yours on.*

Epistemic stance features are used "to present speaker comments on the status of information in a proposition" (Biber et al. 1999:972). Two subcategories are used: modal verbs such as *can*, *could*, *would*, and so on and epistemic adverbs such as *really* and *probably* (cf. Biber et al. 1999:982). In Hallidayan terms, both modal verbs and epistemic adverbs are part of the system of modality whose primary function is "that of assessment: modality construes a region of uncertainty where I can express, or ask you to express, an assessment of the validity of what is being said" (Halliday & Matthiessen 2004).

Affective stance features, finally, are used "to report personal attitudes or feelings" (Biber et al. 1999:974). They are here divided into "affect adjectives" such as *good* or *lovely*,[13] "affect verbs" such as WANT or LOVE, and "primary interjections" such as *oh* or *aye*, which are distinguished from "secondary interjections" such as *boy* or *hell* in that "they are not used otherwise" (Ameka 1992:105).[14] Of interjections, Biber et al. (1999:1083) say that they "have an exclamatory function, expressive of the speaker's emotion."

In total, thirty-one Relation Management features were analyzed; twenty-five (81 percent) are more frequent for women. Among the twenty-one statistically significant features, nineteen (91 percent) are more frequent for women.

As can be seen from Table 6, all the features assembled under politeness are more frequent for women. On the whole, then, the findings support the common perception of women being more polite (e.g., Coates 1993; Holmes 1995). However, interesting nuances can be observed. The differences in the use of *please* are both small and statistically insignificant. This may not be surprising considering that because of its co-occurrence with requests (which are generally seen as threats to negative face) *please* essentially acts as a marker of *negative* politeness while "women are generally more *positively* polite than men" (Holmes 1995:57, emphasis added). It is perhaps more surprising to see that mitigating *though* too is only insignificantly more frequent for women given that the mitigation of disagreement has been found to be a major concern for women (Holmes 1995:64). Significant differences, however, are found in the use of the hedges *I think* (.13), *I don't think* (.14), and, most importantly, *perhaps* (.18) as well as the mitigators *let's* (.10) and *I don't know* (.16). So, all in all, being polite in the sense of protecting face and mitigating the force of face-threatening acts might indeed be a greater concern for women than men.

Table 6. Relation Management Features

Feature		NFm	NFf	Coeff.	χ^2
Politeness					
Hedges	I think	15.31	19.87	.13	**124.77**
	I don't think	3.32	4.40	.14	**31.18**
	I suppose	2.48	2.82	.06	4.47
	perhaps	1.65	2.39	.18	**27.66**
Mitigators	I don't know	7.64	10.55	.16	**96.83**
	though (ADV)	5.82	6.12	.03	1.55
	please	3.83	4.07	.03	1.48
	let's (VERB)	2.65	3.24	.10	**12.18**
Epistemic stance					
Modal verbs	will/'ll	53.54	60.62	.06	**97.75**
	can	44.91	48.80	.04	**35.63**
	would/'d	28.29	34.24	.10	**121.97**
	could	14.82	17.30	.08	**40.63**
	might	6.38	7.88	.11	**32.94**
	shall/'ll	2.98	3.73	.11	**17.31**
Epistemic adverbs	really	13.99	19.72	.17	**204.19**
	actually	6.59	6.08	–.04	4.33
	probably	5.77	5.69	–.01	0.11
Affective stance					
Affect adjectives	good	20.23	20.92	.02	2.47
	nice	7.19	11.40	.23	**196.34**
	bloody	4.67	4.82	.02	0.47
	bad	4.31	4.43	.01	0.32
	lovely	2.23	4.76	.36	**183.28**
Positive affect verbs	WANT (VERB)	24.06	28.79	.09	**90.77**
	LIKE (VERB)	9.83	14.88	.20	**214.06**
	LOVE (VERB)	1.81	3.02	.25	**61.46**
Primary interjections	oh	72.98	91.71	.11	**484.02**
	ah	13.08	10.18	–.12	**79.07**
	ooh	6.63	8.37	.12	**42.10**
	aye	6.42	3.42	–.30	**205.20**
	ha	4.53	4.02	–.06	6.43
	aha	2.39	2.34	–.01	0.09

Note: NFm = normed frequency/1,000 utterances by men; NFf = normed frequency/1,000 utterances by women; coeff. = difference coefficient value; χ^2 = chi-square value ($p < .01$, $df = 1$, critical value = 6.63).

A slightly more mixed picture emerges for epistemic stance features. While all the modal verbs are more frequent for women, among the epistemic stance adverbs only *really* is more frequent for women; *actually* and *probably* by contrast are more frequent for men, although insignificantly. Overall, however, it seems that the "region of uncertainty" referred to above is given more space in women than in men.

The picture is even more mixed for affective stance features. On one hand, all affect adjectives and affect verbs are more frequent for women, with very high rates for

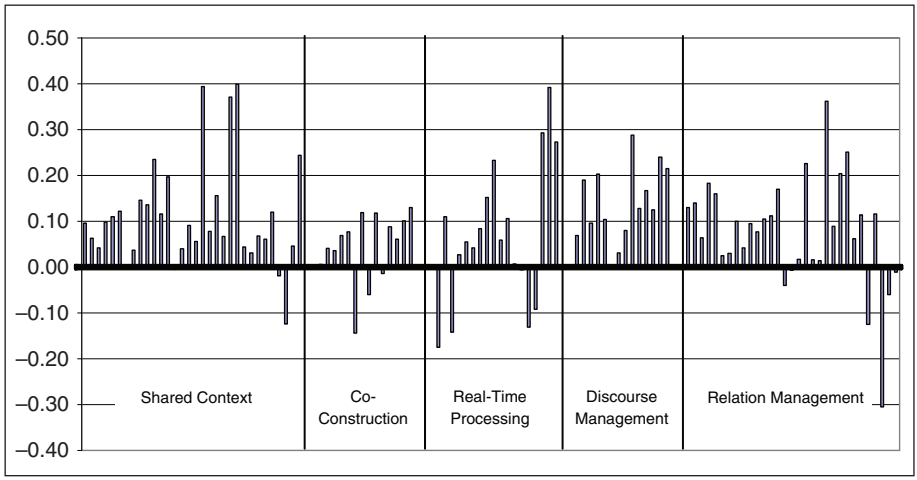

Figure 3. Difference coefficients summary

women for the affect adjectives *nice* (.23) and, particularly, *lovely* (.36; also see Aston & Burnard 1998; Schmid 2003:192) and the affect verbs LIKE (.20) and LOVE (.25). On the other hand, most interjections prove to be more frequent for men: apart from *oh*, which is according to Biber et al. (1999:1083) "by far the most common interjection," and its variant *ooh*, which are more frequent for women, *aha* (however insignificantly), *ha, ah*, and, most prominently, *aye* (–.30) are more frequent for men.

Despite the perhaps surprising figures for interjections, the findings overall suggest that women are more closely oriented toward the goal fundamental to conversation: by using more linguistic features that serve to hedge, protect face, mitigate face threats, allow for uncertainty, and display and create involvement through expressiveness, women seem to show a greater concern for maintaining and establishing rapport.

Discussion

Figure 3 presents all 110 coefficient values as bar charts. It can be seen immediately that the overwhelming majority is upright, representing positive values, which represent those features that are more frequent for women than for men.

Table 7 gives a summary of all results. As shown in Table 7, 93 out of 110 features (84 percent) are more frequent for women. A total of 86 features show significant differences in occurrence between men and women at the $p < .01$ level; 77 of these 86 features (90 percent) are more frequent for women. The statistics are thus unmistakable: they strongly suggest that conversational grammar features are used more frequently by women than by men. Consistently higher female rates occur with regard to a number of subcategories. Table 8 displays all subcategories with at least three members that are invariably more frequent for women.

Table 7. Summary of Results

	Female	Male	Total
Number of features	93	17	110
Percentage	84	16	100
Number of features significant at $p < .01$	77	9	86
Percentage	90	10	100

Table 8. Subcategories Invariably More Frequent for Women

Shared Context	Person deixis
	Time deixis
	Pronouns
Co-Construction	Question tags
Real-Time Processing	Verbal contractions
	Generalization
Discourse Management	Discourse markers
	Quotatives
	Utterance openers
Relation Management	Hedges
	Mitigators
	Modal verbs
	Affect adjectives
	Positive affect verbs

These quantitative findings may be interesting in themselves but they become much more meaningful if we view conversational grammar in the context of its situation: conversational grammar can be appreciated as a set of forms that are useful because they are adapted to constraints set by the situation. Seen against this background, the findings support the feminine grammar hypothesis set out at the beginning of this article, that women's conversational language is more adapted to the conversational situation than men's. I return to this point in the concluding section.

This conclusion seems straightforward, but it is important to take some serious methodological limitations into account. These are briefly discussed in the remainder of this section.

First, the unit of analysis underlying all frequency analyses in this study is frequency for men or women per 1,000 utterances. In adopting this methodology I have followed Aston and Burnard (1998:123) whose analysis of how *lovely* distributes across the sexes in the spoken part of the BNC is based on differences of occurrence per 1,000 utterances for men and women, respectively. However, as noted, utterances by men and women are unequally long: men's utterances are on average 9.53 words long, whereas female utterances are 10.33 words long. So female utterances are on average 0.8 words longer. While this may seem a small difference, it is by no means

negligible. It may safely be assumed that at least some differences may decrease and a few might even disappear altogether as soon as the differing utterance lengths are taken into account. The differences between men and women found in this study, however, are so pronounced and so consistent that even if utterance length were incorporated, these differences seem unlikely to disappear completely.

Second, if the claim is made that women use conversational grammar more frequently than men, one would ideally want to have an exhaustive list of all features distinctive of conversational language. That is impossible to come by because more than any other type of language, conversational language is subject to change. And because many low-frequency distinctive features may not have found their way into a corpus such as the BNC, it would be desirable to have at least a cross-section of distinctive features that is reliably representative of the whole set of linguistic features distinctive of conversation. No such representativeness can be claimed for the set of 110 features examined in this study. Rather, the guiding principle in including these features and excluding others that research has found to be equally distinctive of conversation was that a distinctive feature should have a raw general frequency of above 1,000 occurrences in the conversational subcorpus of the BNC. As a result, a great many features that are strikingly distinctive of conversation, such as the quotatives GO and BE *like* (e.g., Stenström, Andersen, & Hasund 2002) to name only two such features, could not be captured and examined in this study. It is hoped nonetheless that the 110 features that were included, and of which a great many proved statistically significant, are enough to permit at least some generalization.

Third, none of what may be considered key in an account of conversational grammar, namely genuinely syntactic features, has been examined. A large number of studies have concerned themselves with syntactic features such as situational ellipsis and, particularly, headers and tails (Aijmer 1989; Carter & McCarthy 1995; Miller & Weinert 1998). The omission of these features is owed to the fact that the BNC, which is tagged not parsed, simply disallows systematic searches for these features.[15]

Fourth, perhaps the most serious limitation pertains to the fact that the effect of only one variable has been examined, speaker sex. Moreover, there has been no consideration of the effects of same-sex versus mixed-sex talk. It is well known that even the variable sex can be usefully differentiated by taking into account whether women and men talk to members of their own sex or whether men and women talk to each other. Reid (1995), for example, shows that while women use more backchannels in same-sex talk than men do, this difference is neutralized in mixed-sex talk (cf. Schmid 2008).[16] It is further well known that age, class, and region also play crucial roles in how speech is differentiated socially. Barbieri (2007), for example, demonstrates how the parameters age and sex subtly interact in the use of quotatives in American English. Stenström, Andersen, and Hasund (2002) examine London teenage "slanguage" (Stenström, Andersen, and Hasund's cover term for slang, swearing, and vague language) and the effects of age, gender, ethnicity, and socioeconomic background on its use. Rühlemann (2007) shows how *I says* is differentiated in terms of not only speaker sex but also age, class, and, most importantly, region. Many more such studies could

be enumerated. Nonetheless, the tendency discovered in this study for conversational grammar to be more frequent for women than for men and thus for women's conversational language to be more adapted to the conversational situation than men's is so strong that it appears that effects by other social variables can only refine this tendency, not annul it. Note also that the differences between men and women in Rayson, Leech, and Hodges (1997:141) were not only similarly pronounced but also "greater than differences based on age or social group." So speaker sex is likely a decisive, if not *the* decisive, factor in the social differentiation of conversational grammar.

Conclusions

This study addressed the question of what effect speaker sex has on the use of conversational grammar. A pragmatic framework was sketched out that is based on the situational factors constituting the conversational situation and within which conversational language can be seen as adapted to constraints set by the situation. A large number of high-frequency features distinctive of conversation were examined. Building on previous research (e.g., Biber et al. 1999; Leech 2000; Rühlemann 2007), each of them was assigned to one of five situational categories: Shared Context, Co-Construction, Real-Time Processing, Discourse Management, and Relation Management. The frequencies of occurrence of these features in utterances by men and women, respectively, were calculated and compared.

The comparisons reveal a striking distributional skew: the overwhelming majority of the features investigated are more frequent for women than for men, with 84 percent of features more often used by women and with even 90 percent of all statistically significant features more frequent in female conversation.

The statistics are very clear. What is less clear is what we conclude from them. It is paramount to stress that two seemingly possible conclusions would be utterly mistaken. First, to interpret the statistics as suggesting that the distributional skew is *caused* by speaker sex would be mistaken because it would conflate correlation with causation. Investigating the *causes* of women and men talking the way they do in conversation is far beyond the aims of this study. What the statistics suggest is much humbler: they suggest that a (noncausal) correlation holds between speaker sex and specific forms of conversational language and that the correlation can be specified in terms of different degrees of adaptation to constraints set by the conversational situation. Second, the evidence does *not* suggest that conversation was somehow specifically "female territory." We have no evidence to support such a claim; rather, we assume that conversation, understood as "the most common, and, it would appear, the most fundamental condition of 'language use' and discourse" (Schegloff 1979:283), is done by men and women alike. What is significantly different between the sexes is the extent to which they make use of the grammar of conversation, understood as the linguistic features distinctive of that register (see the introduction): female conversationalists seem to deploy these features more extensively than male conversationalists.

The explanation offered here is that in a textual pragmatics in which grammar is regarded as being codetermined by the speech situation (Leech 1983), the grammar of conversation is seen not as independent from the situation that gives rise to conversation as a register but as adapted to it. In this perspective, the findings can usefully be explained with recourse to the feminine grammar hypothesis: according to this hypothesis, women's conversational language is more adapted to the conversational situation than men's. The evidence is strongly in favor of this hypothesis.

As noted in the previous section, however, some limitations surround this interpretation, the most serious being that speaker sex was examined in isolation from other social variables such as age, class, or mixed- versus same-sex conversation. Therefore, some caution is in order here. Before we can establish with confidence that female conversational language exhibits greater adaptation to the conversational situation, more studies are needed that are broader in scope, involving even larger numbers of features, and deeper in analysis, examining not speaker sex alone but also the subtle interplay of effects produced by other core social variables. For the time being though, this study, however limited and initial, does suggest that constraints set by the situation make themselves felt more clearly in female than in male conversational language and that, as a result, the former is more "in line" with the conversational situation than the latter.

This finding ties in well with previous research. First, the greater orientation to context of female language use is a well-established notion in much sociolinguistic research. Holmes (1997:198), for example, reviews relevant research concluding that "women use a wider range of linguistic variants than men, and that their usage varies according to identifiable contextual factors." The "identifiable contextual factor" in whose light language use was examined in this study is the conversational situation (including its factors). However, the "conversational situation" should not be confused with "conversational situations" (in the plural) because, in the Hallidayan sense in which the term is deployed here, the notion of "conversational situation" is an abstraction: "Looking at how people actually use language in daily life, we find that the apparently infinite number of different possible situations represents in reality a very much smaller number of general *types* of situation" (Halliday 1978:29). When talking of the conversational situation we are really talking of the conversational situation *type*: that is, we are talking of those situational factors only (e.g., Shared Context, Co-Construction, etc.) that typically bear on how language is used regardless of the myriad situational differences between all possible situations in which English native speakers (can) converse. The conversational situation, then, is an abstract type of context. When looking at its influence on conversational language use, we can describe this influence only abstractly; we cannot determine the influence that less-than-typical contextual factors undoubtedly have. For example, as noted repeatedly, whether or not conversations are same-sex or mixed-sex conversations is a key contextual variable that will inevitably leave its fingerprint on language use. However, since conversation is not *typically* either between men or between women or between the two together but rather not predetermined with regard to sex composition, this contextual influence

cannot be measured by our model. What we can say, based on the findings of this study, is that we have evidence that women adapt their conversational language to constraints set by the conversational situation understood as an abstract (but nonetheless powerful) contextual dimension more than men do.

Second, despite some evidence to the contrary (for a balanced discussion, see Coates 1993), a widely supported hypothesis holds that women are in the vanguard of linguistic change (e.g., Labov 1990).[17] Since conversation is regarded as the "laboratory for linguistic innovations" (Hughes 2002:15) and since it is in conversation that "the semogenic potential is most likely to get extended" (Halliday 2006:294), it seems consistent to argue that women, being those who exploit the resources provided by conversational grammar more fully, should also be those most likely to expand its boundaries. It is admitted that making such a connection may seem like a stretch because, to my knowledge, none of the features presented here have been identified as particularly innovative in the literature. I wish to make this connection nonetheless on two grounds. First, that none of the features have as yet been shown to be innovations is not proof that they are not innovations but merely reflects the fact that their being innovative or otherwise has not yet been examined. Indeed, some of the features discussed here have hardly been examined at all. For example, I am aware of only one detailed (synchronic) study concerned with the quotative form *I says* (Rühlemann 2007) and only one in-depth (synchronic) study dealing with the quotative form *I goes* (Rühlemann 2008). Given that research into these features has only just started and no diachronic data are available, we cannot rule out the possibility that these features are "new" or spreading from the north of England (*I says*) and from teenage usage (*I goes*) to other dialect areas and other speaker groups, respectively. Second, the cutoff range was set to features with raw frequencies above 1,000 occurrences in the conversational subcorpus of the BNC. Innovative features, being new usage, rarely become that frequent as long as they are new. When they do, they're no longer new. Below the 1,000-occurrence threshold, it is very likely that there are innovations. A case in point is *BE like*, a feature that has attracted a wealth of research because it is regarded as a newcomer to the pool of quotatives. In contrast to, for example, North American English (e.g., Tagliamonte & D'Arcy 2004; Fairon & Singler 2006), where *BE like* has made major inroads, a low frequency was observed for *BE like* in British English in the 1990s: Rühlemann (2007) found a mere fifty-six quotative uses of *BE* immediately followed by *like* (cf. Miller & Weinert 1998; Andersen 2001). However, comparative research by Tagliamonte and Hudson (1999) suggests that *BE like* may be gaining in frequency there as well. Crucially, in the present connection, to judge by the BNC, *BE like* is strikingly female in distribution (note that the spoken data in the BNC stem from the early 1990s): a difference coefficient of +.714 was obtained in Rühlemann (2007:219). So although admittedly *BE like* is just one innovative feature and the numbers of attested uses in the BNC are relatively low, it seems justifiable, to an extent, to hypothesize that one of the reasons that women are frequently in the vanguard of language change may be that women exploit the linguistic resources of conversational grammar more fully than men do. I leave this hypothesis to be tested for future research.

To conclude, a strictly tentative note should be added relating to conversational grammar as a resource for doing gender and its future development. Gender norms are not stable across time but in flux. Individuals not only replicate but also can create, challenge, and subvert sociolinguistic norms (Holmes 1997:196). Moreover, changes in society as a whole may contribute to changes in the perception of what is gender-appropriate linguistic behavior. For example, in media discourse—no doubt an important discourse type of modern life—a trend has been observed toward conversationalization (Fairclough 1995), that is, a process whereby linguistic devices, styles, and, one might add, the grammar typical of conversation are used to make public language resemble the language of ordinary conversation. It might be possible that, as a consequence of such conversationalization, conversational grammar, hitherto firmly associated with the conversational language of women, will lose (some of) this association because it may be making inroads into the conversational language of men. If that happens, it will be clear proof that the use of conversational grammar is gender, not sex, specific. Whether it will happen is still an open question. Addressing this question will be greatly facilitated if and when a corpus such as the twentieth-century BNC will be complemented by a comparable corpus that captures the conversational language use by men and women in the early twenty-first century.

Acknowledgments

I'm greatly indebted to the editors, Anne Curzan and Robin Queen, as well as two anonymous reviewers whose comments on earlier versions of this article were an invaluable help. Whatever faults remain must be laid entirely at my door.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The author(s) received no financial support for the research and/or authorship of this article.

Notes

1. In the *Longman Grammar of Spoken and Written English*, *I* and *you*, respectively, are nineteen and thirty times as frequent in conversation as in academic writing (Biber et al. 1999:334).
2. As an anonymous reviewer noted, social norms are "cyclical" phenomena in that norms, because they are norms, are likely to be adhered to by the majority of speakers, who thus reinforce the norms.
3. The clearest example of endophoric (rather than exophoric) deixis is what is called discourse deixis as realized by expressions that refer to (portions of) the discourse itself (e.g., discourse markers); some also view certain types of cataphora as endophoric deictic markers.
4. As an anonymous reviewer noted, the tendency to build same-sex networks may be not so much human as specific to societies that are highly gender segregated.

5. All illustrative examples in this study are taken from the conversational subcorpus of the British National Corpus (BNC).
6. The annotation of the BNC captures a wealth of paralinguistic events. These include laughter in two forms: laughter as a voice quality (within-speech laughter), as in *[laughing] Our Dawn thinks she's at a tea party []!* where the laughing voice quality "colors" the whole utterance, and laughter as a vocal event (between-speech laughter), as in *[laugh] . . . You need a bib mother*, where laughter is separated from the adjacent words. This latter type of laughter can stand alone and is often used as a backchannel form (see the section on Co-Construction features).
7. In addition to indexing humor, laughter may fulfill important discourse functions. Günther (2003), for example, proposes a taxonomy of discourse functions of laughter including not only "affiliative laughter" (most typically in response to humor) but also, inter alia, "disaffiliative laughter," which "expresses disapproval or criticism" (Günther 2003:156; also see the section on Co-Construction below).
8. Following common practice in corpus linguistics, headwords (lemmas), such as BE in Table 3, are indicated by small capitals.
9. Possessive *me* is a good example showing that situational factors, such as, in this case, Real-Time Processing, are not the only constraints influencing the use of conversational grammar. As Hollmann and Siewierska (2007:413) demonstrate, use of possessive *me* is also phraseologically constrained in that it most commonly co-occurs with nouns that have an "inalienable" semantics, such as kinship terms (*me mum*) and body parts (*me back*). To complicate matters, it should also be noted that possessive me may well be a feature of regional dialects (cf. Anderwald 2004). If it is part of a speaker's dialect, it is as yet unclear whether phraseology and/or Real-Time Processing still play any role or whether it is simply a systematic part of the speaker's grammar.
10. There are, of course, Standard English uses of *I were*, for example, in *if*-clauses as in *Oh I should make it if I were you*; inspection of concordance lines in the conversational subcorpus, however, suggests that these standard uses are by far outnumbered by nonstandard uses.
11. Both *I says* and *I goes* are seen as adapted to Real-Time Processing constraints on the following grounds (see Rühlemann 2007, 2008). Both forms are "multiturn quotatives," that is, they occur preferably in presentations of extended dialogue with frequent turn taking and, therefore, most often alternate with the forms *he*/*she says* and *he*/*she goes*, respectively. Instead of marking the switch from third person to first person morphologically and phonologically, which increases complexity and processing cost, by using the Standard English forms *say* and *go*, respectively, the third-person forms *says* and *goes* are used throughout the presentation of the anterior conversation, thereby not only achieving a symmetry of forms but also reducing processing cost by reducing complexity.
12. The distinction between hedges and modality devices is admittedly a difficult one: a hedge such as *I think* may convey not only the speaker's wish to respect the addressee's autonomy but also the speaker's uncertainty with regard to the proposition; conversely, a modal verb such as *would* may be used for politeness purposes, as in *Mike come out would you mind making a drink?*
13. As Biber et al. (1999:516) observe, *good, bad, lovely,* and *nice* are the most frequent predicative adjectives in conversation; *bloody*, by contrast, is typically used as an intensifier, as

in *You're stupid, you're bloody stupid!* (Biber et al. 1999:564). Grammatically, the classification of *bloody* as an adjective is hence questionable.

14. Because secondary interjections are used "otherwise," as Ameka notes, they cannot exhaustively be searched for in a part-of-speech-tagged corpus such as the BNC and are therefore not included in this study.
15. To my knowledge, even parsed (i.e., syntactically annotated) corpora, such as the International Corpus of English (ICE) corpora, are not marked up for situational ellipsis, headers, tails, or any other specifically conversational type of syntax.
16. Interestingly, Schmid (2008), working on data from the ICE-GB, found that the effect of neutralization of differences in mixed-sex conversations was largely due to accommodation by men.
17. For a very recent counterexample, see Torgersen and Gabrielatos (2009), in whose Linguistic Innovators Corpus use of the innovative tag *you get me* is led by inner-city, non-Anglo males.

References

Adolphs, Svenia. 2008. *Corpus and context: Investigating pragmatic functions in spoken discourse*. Amsterdam: John Benjamins.

Adolphs, Svenia & Ron Carter. 2003. *And she's like it's terrible, like*: Spoken discourse, grammar, and corpus analysis. *International Journal of English Studies* 3(1). 45-56.

Aijmer, Karin. 1989. Themes and tails: the discourse functions of dislocated elements. *Nordic Journal of Linguistics* 12. 137-154.

Allwood, Jens, Joakim Nivre & Elisabeth Ahlsén 1990. Speech management: On the non-written life of speech. *Nordic Journal of Linguistics* 13. 3-48.

Ameka, Felix. 1992. Interjections: The universal yet neglected part of speech. *Journal of Pragmatics* 18. 101-118.

Andersen, Gisle. 1998. The pragmatic marker *like* from a relevance-theoretic perspective. In Andreas H. Jucker & Yael Ziv (eds.), *Discourse markers: Descriptions and theory*, 147-170. Amsterdam: John Benjamins.

Andersen, Gisle. 2001. *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. Amsterdam: John Benjamins.

Anderwald, Lieselotte. 2004. The varieties of English spoken in the southeast of England: Morphology and syntax. In Bernd Kortmann & Edgar Werner Schneider (eds.), *A handbook of varieties of English: A multimedia reference tool*, vol. 2, 175-195. Berlin: Mouton de Gruyter.

Arndt, Horst & Richard W. Janney. 1987. *InterGrammar: Toward an integrative model of verbal, prosodic and kinesic choices in speech*. Berlin: Mouton de Gruyter.

Arndt, Horst & Richard W. Janney. 1991. Verbal, prosodic, and kinesic emotive contrasts in speech. *Journal of Pragmatics* 15. 521-549.

Aston, Guy & Lou Burnard. 1998. *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh, UK: Edinburgh University Press.

Barbieri, Federica. 2007. Older men and younger women: A corpus-based study of quotative use in American English. *English World-Wide* 28(1). 23-45.

Barbieri, Federica. 2008. Patterns of age-based linguistic variation in American English. *Journal of Sociolinguistics* 12(1). 58-88.

Biber, Douglas & Edward Finegan. 1988. Adverbial stance types in English. *Discourse Processes* 11. 1-34.

Biber, Douglas & Edward Finegan. 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text* 9(1). 93-124.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow, UK: Pearson.

BNC Consortium, Oxford University Computing Services. 2001. *The British National Corpus version 2* (BNC World). http://www.natcorp.ox.ac.uk (2 July 2009).

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language use*. Cambridge, UK: Cambridge University Press.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37(4). 575-596.

Carter, Ron & Michael J. McCarthy. 1995. Grammar and the spoken language. *Applied Linguistics* 16(2). 141-158.

Carter, Ron & Michael J. McCarthy. 2006. *Cambridge grammar of English*. Cambridge, UK: Cambridge University Press.

Coates, Jennifer. 1993. *Women, men and language. A sociolinguistic account of sex differences in language*, 2nd edn. Harlow, UK: Longman.

Coates, Jennifer. 2006. *"It was so funny": Gender and humour in everyday talk*. Harrogate, UK: 40th IATEFL Annual Conference paper.

Conrad. Susan. 2004. Corpus linguistics, language variation, and language teaching. In John M. Sinclair (ed.), *How to use corpora in language teaching*, 67-85. Amsterdam: John Benjamins.

Crowdy, Steve. 1994. Spoken corpus transcription. *Literary and Linguistic Computing* 9(1). 25-28.

Diani, Guliana. 2004. The discourse functions of *I don't know* in English conversation. In Karin Aijmer & Anna-Brita Stenström (eds.), *Discourse patterns in spoken and written corpora*, 157-171. Amsterdam: John Benjamins.

Fairclough, Norman. 1995. *Media discourse*. London: Arnold.

Fairon, Cédrick & John V. Singler. 2006. I'm like, "Hey, it works!": Using GlossaNet to find attestations of the quotative *(be) like* in English-language newspapers. In Antoinette J. Renouf & Andrew Kehoe (eds.), *The changing face of corpus linguistics*, 325-336. Amsterdam: Rodopi.

Ferrara, Kathleen & Barbara Bell. 1995. Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *be + like*. *American Speech* 70(3). 265-290.

Fraser, Bruce. 1990. An approach to discourse markers. *Journal of Pragmatics* 14. 383-395.

Goffman, Ervin. 1981. *Forms of talk*. Philadelphia: University of Philadelphia Press.

Grice, Paul H. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics III*, 43-58. New York: Academic Press.

Günther, Ulrike. 2003. *What's in a laugh? Humour, jokes and laughter in the conversational subcorpus of the BNC*. Freiburg, Germany: University of Freiburg dissertation. http://www.freidok.uni-freiburg.de/volltexte/735/ (2 July 2009).

Halliday, M. A. K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. 2006. Afterwords. In Geoff Thompson & Susan Hunston (eds.), *System and corpus: Exploring connections*, 293-299. London: Equinox.

Halliday, M. A. K. & Christian M. I. M. Matthiessen. 2004. *An introduction to functional grammar*, 3rd edn. London: Edward Arnold.

Helasvuo, Marja-Liisa. 2004. Shared syntax: The grammar of co-constructions. *Journal of Pragmatics* 36. 1315-1336.

Hollmann, Willem & Anna Siewierska. 2007. A construction grammar account of possessive constructions in Lancashire dialect: Some advantages and challenges. *English Language and Linguistics* 11(2). 407-424.

Holmes, Janet. 1995. *Women, men and politeness*. Harlow, UK: Longman.

Holmes, Janet. 1997. Women, language and identity. *Journal of Sociolinguistics* 1(2). 195-223.

Holmes, Janet & Maria Stubbe. 1997. Good listeners: Gender differences in New Zealand conversation. *Women and Language* 20(2). 7-14.

Hughes, Rebecca. 2002. *Teaching and researching speaking.* Harlow, UK: Pearson.

Hunston, Susan & Geoff Thompson. (eds.) 2000. *Evaluation in text: Authorial stance and the construction of discourse.* Oxford: Oxford University Press.

Johnstone, Barbara. 1993. Community and contest: Midwestern men and women creating their worlds in conversational storytelling. In Deborah Tannen (ed.), *Gender and conversational interaction*, 63-80. New York: Oxford University Press.

Kaltenböck, Gunther. 2004. Using non-extraposition in spoken and written texts. In Karin Aijmer & Anna-Brita Stenström (eds.), *Discourse patterns in spoken and written corpora*, 219-242. Amsterdam: John Benjamins.

Kilgarriff, Adam. 1998. *BNC database and word frequency lists*. http://www.kilgarriff.co.uk/bnc-readme.html (2 July 2009).

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2. 205-254.

Leech, Geoffrey. 1983. *Principles of pragmatics*. London: Longman.

Leech, Geoffrey. 1999. The distribution and function of vocatives in American and British English. In Hilde Hasselgard & Signe Oksefjell (eds.), *Out of corpora: Studies in honour of Stig Johansson*, 107-118. Amsterdam: Rodopi.

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

Leech, Geoffrey & Roger Fallon. 1992. Computer corpora—What do they tell us about culture? *ICAME Journal* 16. 29-50.

Lerner, Gene. H. 1991. On the syntax of sentences-in-progress. *Language in Society* 20. 441-458.

Malinowski, Bronislaw. 1923. The problem of meaning in primitive languages. In Charles K. Ogden & Ivor A. Richards (eds.), *The meaning of meaning*, 296-336. London: Routledge.

McCarthy, Michael J. & Ron Carter. 2004. "There's millions of them": Hyperbole in everyday conversation. *Journal of Pragmatics* 36. 149-184.

McCarthy, Michael J. & Anne O'Keeffe. 2003. What's in a name? Vocatives in casual conversation and radio-phone-in calls. In Pepi Leistyna & Charles F. Meyer (eds.), *Corpus analysis: Language structure and language use*, 153-185. Amsterdam: Rodopi.

McEnery, Tony & Andrew Wilson. 1996. *Corpus linguistics*. Edinburgh, UK: Edinburgh University Press.

McEnery, Tony & Richard Xiao. 2004. Swearing in modern British English: The case of *fuck* in the BNC. *Language and Literature* 13(3). 235-268.

Miller, Jim & Regina Weinert. 1998. *Spontaneous spoken language: Syntax and discourse*. Oxford, UK: Clarendon.

O'Keeffe, Anne. 2004. "Like the wise virgin and all that jazz": Using a corpus to examine vague categorisation and shared knowledge. In Ulla Connor & Thomas A. Upton (eds.), *Applied corpus linguistics: A multidimensional perspective*, 1-20. Amsterdam: Rodopi.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Rayson, Paul, Geoffrey Leech & Mary Hodges. 1997. Social differentiation in the use of English vocabulary: Some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics* 2(1). 133-152.

Reid, Julie. 1995. A study of gender differences in minimal responses. *Journal of Pragmatics* 24. 489-512.

Romaine, Suzanne & Deborah Lange. 1991. The use of *like* as a marker of reported speech and thought: A case of ongoing grammaticalization in progress. *American Speech* 66(3). 227-279.

Rühlemann, Christoph. 2007. *Conversation in context: A corpus-driven approach*. London: Continuum.

Rühlemann, Christoph. 2008. Conversational grammar–bad grammar? A situation-based description of quotative *I goes* in the BNC. *ICAME Journal* 32. 157-177.

Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4). 696-735.

Schegloff, E. A. 1979. The relevance of repair for syntax-in-conversation. In Talmy Givón (ed.), *Discourse syntax*, 261-286. New York: Academic Press.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge, UK: Cambridge University Press.

Schleppegrell, Mary J. 1991. Paratactic *because*. *Journal of Pragmatics* 16. 323-337.

Schmid, Hans-Jörg. 2003. Do women and men really live in different cultures? Evidence from the BNC. In Andrew Wilson, Paul Rayson & Tony McEnery (eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech*, 185-221. Frankfurt, Germany: Peter Lang.

Schmid, Hans-Jörg. 2008. *Investigating style. Lexico-grammatical corpus evidence for the socio-situational construction of discourse identities*. Freiburg, Germany: First Conference of the International Society of the Linguistics of English paper.

Stenström, Anna-Brita. 1998. From sentence to discourse: *Cos (because)* in teenage talk. In Andreas H. Jucker & Yael Ziv (eds.), *Discourse markers: Description and theory*, 127-146. Amsterdam: John Benjamins.

Stenström, Anna-Brita, Gisle Andersen & Ingrid K. Hasund. 2002. *Trends in teenage talk*. Amsterdam: John Benjamins.

Stenström, Anna-Brita & Jan Svartvik. 1994. Imparsable speech: Repeats and other nonfluencies in spoken English. In Nelleke Oostdijk & Pieter de Hahn (eds.), *Corpus-based research into language: In honour of Jan Aarts*, 241-254. Amsterdam: Rodopi.

Tagliamonte, Sali & Alex D'Arcy. 2004. *He's like, she's like*: The quotative system in Canadian youth. *Journal of Sociolinguistics* 8(4). 493-514.

Tagliamonte, Sali & Rachel Hudson. 1999. *Be like* et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3(2). 147-172.

Tannen, Deborah. 1990. *You just don't understand*. New York: William Morrow.

Thompson, Geoff & Susan Hunston. (eds.). 2006. *System and corpus: Exploring connections*. London: Equinox.

Torgersen, Eivind & Costas Gabrielatos. 2009. *A corpus-based study of invariant tags in London English*. http://eprints.lancs.ac.uk/26026 (2 July 2009).

Tottie, Gunnel. 1991. *Negation in English speech and writing: A study in variation*. San Diego, CA: Academic Press.

Tottie, Gunnel & Sebastian Hoffmann. 2006. Tag questions in British and American English. *Journal of English Linguistics* 34. 283-311.

West, Candace & Don H. Zimmermann. 1987. Doing gender. *Gender & Society* 1(2). 125-151.

Wichmann, Anne. 2004. The intonation of *please*-requests: A corpus-based study. *Journal of Pragmatics* 36. 1521-1549.

Wong, Deanna. 2008. *A corpus based analysis of the pragmatic dimensions of backchannels in Australian English telephone conversations*. Ascona, Switzerland: 29th ICAME conference paper.

Wong, Deanna & Pam Peters. 2007. A study of backchannels in regional varieties of English, using corpus mark-up as means of identification. *International Journal of Corpus Linguistics* 12(4). 479-509.

Xiao, Richard & Hongyin Tao. 2007. A corpus-based sociolinguistic study of amplifiers in British English. *Sociolinguistic Studies* 1(2). 241-273.

Bio

Christoph Rühlemann gives lectures as the modern English linguistics chair at Ludwig-Maximilians-University in Munich, Germany.