1-1-2004

Lagged Co-Occurrence Analysis reveals gender differences of co-occurrence patterns in personal narratives

Shuki Cohen

Lagged Co-Occurrence Analysis reveals gender differences of co-occurrence patterns in personal narratives

Lagged Co-occurrence Analysis Reveals Gender Differences of Co-occurence Patterns in Personal Narratives[*]

Shuki Cohen

1 Introduction

Gender difference has long been the focus of extensive psychological as well as linguistic research. Differences in behavior, attitudes, symptoms and other psychological and sociological characteristics have been extensively observed and documented (for reviews see e.g. Kessler, 1994; Brannon, 2002). Viewing language as one such behavior, gender differences in naturally-occurring speech have long been discussed in the linguistic research literature (for reviews see e.g. Haas, 1979; Mulac and Lundell, 1986). Gender difference in language use and comprehension has even become a popular topic of discussion in the non-professional readership as well (Tannen, 1991, 1994; Pease and Pease, 2001).

Most empirical research examining the differences in language use across the gender line can be categorized into 2 major divisions: form (including style and grammar), and content, usually practiced as content analysis. Research on the different speech style that characterizes women's speech has found differences in the way in which women convey authority (e.g. Levenston, 1969; Tannen, 1994; Mott and Petrie, 1995), uncertainty (e.g. Lakoff, 1975, 1977; McMillan, 1977; Mulac et al., 1998), politeness (e.g. Brown and Levinson, 1987; Holmes, 1995) and disagreements (e.g. Carli, 1990; McLachlan, 1991), as well as other complex pragmatic tasks.

Content analysis studies, on the other hand, rely on systematic categorization of the words uttered by the informant. This is usually performed either using computer software (e.g. Stone, 1966), or trained raters (e.g. Gottschalk, 1969). Only a few studies have quantitatively examined gender differences through the content analysis of male and female speakers (e.g. Gleser et al., 1959; Mulac and Lundell, 1986; Ries, 1999). These studies have consistently found that women tend to use more emotional words (e.g. Mulac et al., 1990; Anderson and Leaper, 1998), adverbs (Mondorf, 2002;

for the particular case of intensive adverbs see e.g. McMillan, 1977; Mulac and Lundell, 1986), hedges and questions (Crosby and Nyquist, 1977; Fishman, 1990; Holmes, 1990; Macaulay, 2001), including tag questions (for review see Haas, 1979; Thorne et al., 1983; Cameron, 1989), among other features.

Research on both the form as well as the content of female and male speech has mostly established the existence of "genderlect", or the preferential use of linguistic features by one gender over the other (Lakoff, 1975; Mulac et al., 2001). However, most of the findings in the studies mentioned so far were shown to be sensitive to a plethora of factors, most notably socio-economic status (Labov, 1966, 1990), the gender composition of the dyads (e.g. Mulac et al., 1988; Bilous and Krauss, 1988), level of familiarity of informants (McLachlan, 1991; Fitzpatrick et al., 1995), status differential between the speakers (e.g. Mott and Petrie, 1995), and more. Some of the variability in the adherence of a subject to their linguistic gender role can be explained by the Speech Accommodation theory, which claims that speakers tend to coordinate their speech with one another (Giles, 1987). This process may be the reason why, in general, genderlects tend to be less salient in mixed-gender dyads than in same-gender dyads. (Bilous and Krauss, 1988; Mulac et al., 1988; Thomson and Murachver, 2001).

Several methodological maneuvers were employed to control for the accommodation process in studying gender-preferential language. The most common one is to compare a setting conducive to high accommodation to one where accommodation is presumably minimal. Experimental designs presumed to confer high pull for accommodation are ones in which there is some discrepancy between participants in certain aspects of their "community of practice" (Cappella, 1994). In these cases, accommodation theory predicts an attenuation of the stylistic markers that typify the individual groups. Thus, the language of female informants in mixed-gender groups does not adhere as closely to the "female register" as their speech in same-gender groups (McMillan et al., 1977; Bilous and Krauss, 1988).

Another approach that minimizes the accommodation effect is the use of monologues or narratives, elicited using only minimal interaction with the subject. This design can be found in psychological settings (Gleser, 1959; Pennebaker, 1995), sociolinguistic settings (Labov, 1990, 1997; Mulac and Lundell, 1986), and public speeches (e.g. Mulac et al., 1986b). It is also widely used in psychodynamic psychotherapy, where (at least in the initial stages) the verbal contribution of the therapist is minimal.

This study has used recorded narratives of an interpersonal nature that were elicited using a silent "confessor", who interacted with the subject minimally before the recording, and did not intervene throughout the sub-

ject's story-telling. Unlike previous studies, the present study is concerned with differences in the co-occurrence patterns that distinguish male from female speech, rather than usage frequency of linguistic features. The premise of this study is that tokens or linguistic markers can occur in the same frequency in male and female speech, and yet have different pattern of occurrence, as revealed by computerized Lagged Co-occurrence Analysis (LCA). These patterns reflect the temporal co-occurrence of the markers at hand within each person's narrative. For example, LCA may show that females tend to utter a certain token in bursts whereas males utter the same token constantly throughout the narrative, even though the overall frequency of utterance for this token may be equal among males and females. LCA calculates the probability that a certain marker will be uttered at any particular lag after the same marker has been uttered. By extension, the technique can also be applied to estimate the probability that two different markers will be separated by a particular lag.

2 Method

Two hundred undergraduates from a large, urban northeastern university were recruited for the study. The students were all enrolled in an introductory course in psychology and participated in the study for course credit. The sample comprised of 155 women and 45 men. Around 60% were Caucasians, 12% Asians, 8% of Spanish descent, 5% African-Americans, 5% Indian and 10% "other" or mixed ethnicity. All subjects were native speakers of North American English and were either born in the US or immigrated to the US before they were 8 years of age. Each subject was asked to tell in their own words the details of a recent disagreement they had with somebody who is emotionally close to them. The subjects were told in advance that they have five minutes to talk and that they should make an effort to speak for the entire time, without the help or the lead of the experimenter. The narratives were recorded and transcribed using a transcription manual popular in the field of psychological transcriptions (Stinson and Mergenthaler, 1992). Each narrative was processed in 3 consecutive stages of quality-check, performed by 3 different associates (all of whom native speakers of North American English) who were trained in this procedure.

Lagged Co-occurrence Analysis (LCA) was performed using software developed by the author. The software converts each narrative into a set of binary series for each target word of interest. For each target word the series assumes the value of 1 in positions in which any subsequent word matches the target word, or 0 if there is no match. Once the series is constructed, it is summarized by a calculation of the overall probability that a target word will

appear in each lag after it was first uttered. The results across all possible lags are presented in the form of a graph (called a "correlogram"), depicting the probability that a target word will be followed by its recurrence for each and every lag after or before it was uttered. Lags are measured in tokens, and so a significant peak when the lag equals 2 connotes a statistically significant likelihood of the target to be uttered again two tokens after it was first uttered.

As mentioned before, the two target words can be different, in which case the algorithm calculates the cross-correlogram. For example, if the narrative is replete with the phrase "I don't know", the cross-correlogram of "I" (here serving as a target) and "know" (here serving as token2) will show a statistically significant peak at lag=+2. The likelihood of the word "know" to appear 2 tokens downstream from "I" will then achieve statistical significance due to the prevalence of the phrase in the narrative.

Detecting statistically significant patterns in the auto- or cross-correlogram rests on the premise that the theoretical cross-correlation between two random series (i.e. series in which the words don't co-occur consistently at any lag) is a flat horizontal line. Any statistically significant deviation from that horizontal line can be interpreted as a sign of co-occurrence of two words at a particular lag in the case of the cross-correlogram, or the word with itself in the case of the auto-correlogram.

For simplicity's sake, only auto-correlation results will be presented in the present study. After computing the auto-correlation for each individual, an average auto-correlogram was obtained for males and females. Any significant difference between the two average correlograms was taken to signify a different pattern of uttering the words under examination between the genders.

Two types of patterns were sought in this study: marker-level and thematic-level. Marker-level patterns include the auto-correlograms of a specific word, such as "I". Thematic-level patterns make use of thematic dictionaries, and look for co-occurrence of words of the same thematic category. The pattern of uttering an emotional word is a relevant example. In this type of analysis, all emotional words in the dictionary are taken as being of a single semantic category, and are therefore equivalent and interchangeable as far as the statistical analysis is concerned. The thematic dictionaries in this study were borrowed from Linguistic Inquiry and Word Count (LIWC, Pennebaker and Francis, 1999). These thematic dictionaries were found to be psychologically meaningful in several studies (for review see Pennebaker and Francis, 1999).

3 Results

3.1 Marker-Level Analysis

In order to establish speech patterns in the utterance of individual markers, psychologically meaningful words were analyzed for patterns in their auto-correlogram. One such candidate word is "I", which was found previously to be used preferentially by males in two studies involving written samples (Mulac et al., 1990; Mulac and Lundell, 1994), but not in a study involving monologues (Raskin and Shaw, 1988). Content-analysis of the marker in our corpus did not show any statistically significant difference between its use by males and females. The mean proportion of the word "I" in males narratives was 0.048 (N=72), and that of women was 0.051 (N=221). Two-tailed t-test (under the assumption of unequal variance) did not reveal any statistically significant difference between males and females in their proportion of use of "I" (t=-1.493; p<0.14; df=292).

However, an examination of the auto-correlogram, based on the occurrences of the word "I", showed distinct dissimilarities in the temporal pattern in which this word was uttered by men and by women. Female subjects avoid uttering the word "I" in the first, second and third position after it was first uttered, while males have only a slight dispreference for uttering an "I" in the second position, and in the third and fourth position exhibit an increased likelihood of uttering another "I". Thus, the auto-correlogram for males and females will show a trough in lag 1. This trough will be deeper for the females, denoting lower likelihood to utter the word. For males, the auto-correlogram shows a much smaller (albeit significant) trough in lag 1, while in positions 2 and 3 there is some evidence for a peak in the correlation, signifying an elevated likelihood of saying "I" 2 or 3 words downstream after it was first uttered. Figure 1 in the next page shows the average auto-correlogram for both males and females.

To ensure the consistency of the pattern for each and every subject, a population distribution of the trough was constructed. This was done by subtracting the average correlation coefficients of lags 1 to 4 (roughly the area of the trough for most subjects) from the background level of lags 10-15 (in which the correlation level is considered incidental as explained above) for each subject. As can be seen in the distribution in Figure 2, most subjects (70%) exhibited a trough in their auto-correlogram, while some (mainly males) exhibited a peak in lags 2-4, as explained previously.

After establishing the consistency of the speech pattern, as well as its different shape for men and women, a narrative excerpt from a "prototypical" exemplar of each gender is presented in the next page.
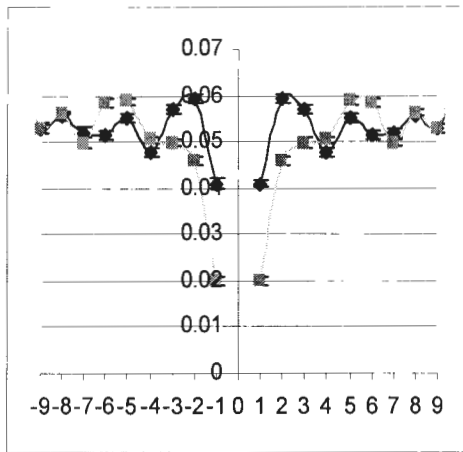
Figure 1: The auto-correlogram of the word "I" in narratives of males and females. The correlation coefficient is given on the y-axis and the lag in words is given in the x-axis. Females are represented by the gray squares and males in the black triangles. The error bars represent the standard error of the mean for the correlation coefficient for each lag. Time 0 was eliminated from the graph as it always equals 1 by definition.
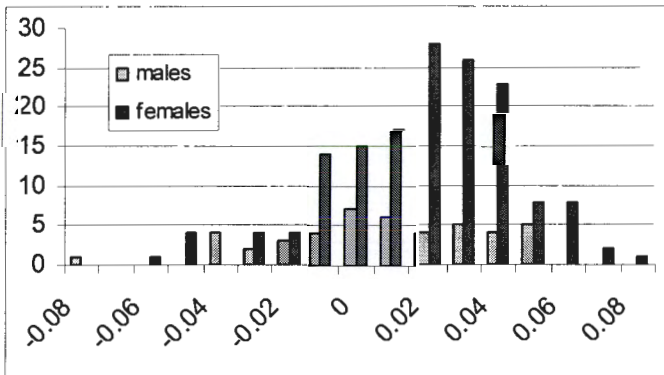

Figure 2: Distribution of the average decrease in the likelihood of saying "I" 1 to 4 words after it was first uttered. The y-axis represents the number of subjects, while the x-axis represents the average correlation in the range of 1-4 words downstream.

In the following excerpt of a "prototypical" female narrative, the word "I" appears in bold and underlined typeface to facilitate the perception of its

temporal structure. The auto-correlation of this subject (computed for the entire narrative) is presented in Figure 3 following the excerpt. Note the complete absence of any subsequent "I" within 2 words after an "I" was first uttered:

"**I** can't just get up and leave, especially after knowing what's required of me, **I** was trying to explain to her that **I** can't do that even though **I** am very popular, people like me at the job, **I** still have to show a good example as a manager in training, so **I**'m not going (snap) to be able to do that. that brought up other issues, **I**'m lying, this that and the other, brought up issues about boyfriends, and **I** didn't even see why that had anything to do with planning a trip to *Miami_Florida. and **I** didn't understand why they couldn't just go without me because they can do that. it's not like **I**'m never going to see them again. **I** know they'll come back, **I** can spend as much time with them when they come back, it's not a big deal, and me going to *Miami_Florida is not a priority for me right now. **I** have bills to pay, **I** have rent to pay so **I** just really, you know can't just forget my job for like a week or so all because of spring break, it's not really a big deal for me".
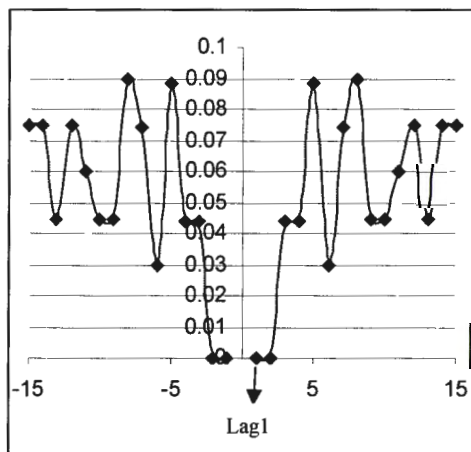

Figure 3: The auto-correlogram associated with the female subject whose speech excerpt is presented in the example above. Note the complete lack of any "I I", "I X I" structures in lags 1 and 2 of the auto-correlation.

Similarly, a "prototypical" male narrative can be demonstrated using the following excerpt:

"**I** thought like he was choking me and that's not like, um **I** didn't think that was fair either. so uh he, **I** fishhooked him and like **I** pulled real hard and he let go and he was like, he, he, he like stood up and right now like **I**, **I** guess **I** have a pretty bad temper and **I**, my uh adrenaline was going pretty well now. and we stood like face to face and he was uh 'oh that was so cheap.' and like **I** ca-, **I** was like catching my breath first because uh (sniff) **I** still had like, - still like loss of breath and like, (sniff) so uh, so he said that was cheap and then uh, and then like, and **I** go, **I**, **I** really just, **I** just like, - **I** didn't say anything and **I** walked into the room and **I** was like real mad and so **I** just sat down and started like typing on my computer…"

The auto-correlogram associated with this subject (computed for the whole narrative) is presented Figure 4 in the next page. Note how "autocorrelated" is this subject's use of the word "I" – after uttering the word, the likelihood of uttering it within 1 or 3 words is greater than chance level. This is due to both false starts (e.g. "I, I really just, I just like") as well as grammatical choices (e.g. "I guess I have a pretty bad temper").
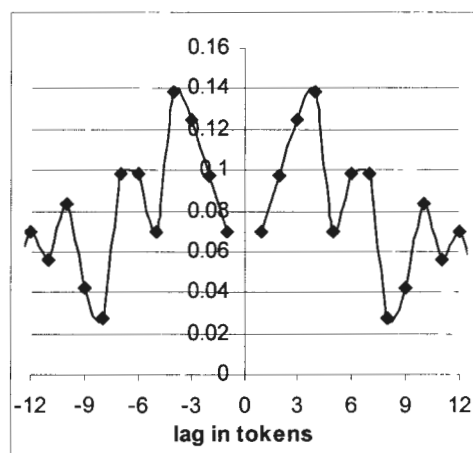

Figure 4: Auto-correlation of a "prototypical" male subject. A speech sample of this narrative is given in the previous page. Note the higher variability in the correlation. This is largely a result of a shorter narrative than the female example, which led to higher variability, due to the smaller sample size.

3.2 Thematic-level analysis

Similar to the case of single words, LCA was used to track the temporal patterns in the utterance of words presumed to be interchangeable, such as words that belong to the same semantic category. One such semantic category that bears psychological significance for gender studies is the family of emotion words.

Gender differences were found in the temporal patterns in which men and women utter emotional words. Figure 5 in the next page shows the auto-correlogram of positive emotion words for male and female subjects. As can be seen in the graph, both men and women refrain from saying another positive emotion word once the first one was uttered. This may well be a result of syntactic constraints common for both men and women. However, men tend to avoid saying emotional words to a larger extent and over a higher number of words downstream. This can be inferred from the deeper and wider trough in the auto-correlogram. Similar to the case with "I" that was discussed above, there was no significant difference in the proportion of positive emotion words in men's and women's narratives. Men's proportion of positive emotion words was 0.016 (N=75) and that of women was 0.014 (N=224), and the difference was not statistically significant (t=1.28; p>0.2; df=298; unequal variance assumed).
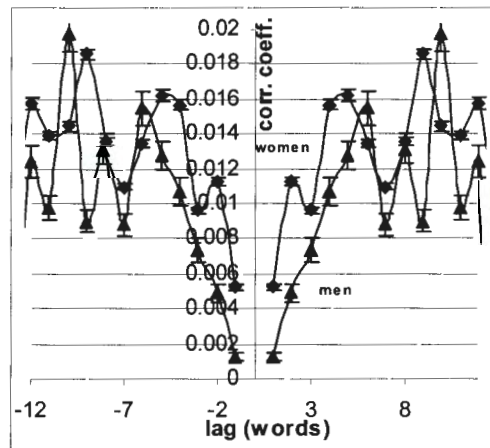

Figure 5: Auto-correlogram of positive emotion words for male and female narratives. Note how the error bars of men are larger than those of women; this is due to the fact that there are more women in the sample combined with the fact that the occurrence frequency of positive emotion words in spontaneous narratives is not high.

4 Discussion

Lagged-Correlation Analysis (LCA) is a powerful tool to detect speech patterns in spontaneous narratives. Further, its accuracy furnishes a sensitive tool for the detection of consistent differences in the temporal pattern with which male and female subjects uttered relevant markers, even when the actual frequencies of those markers were similar in male and female narratives.

Several factors seem to govern the shape of the lagged-correlation. These can roughly be categorized into the syntactic/stylistic factor, the psychological factor and the statistical factor.

The syntactic/stylistic factor seems to govern the overall shape of the LCA correlogram. Whenever syntactic or stylistic constraints prohibit the co-occurrence of certain markers in succession or close proximity the correlogram will most likely have a trough around the first few positions after the marker was uttered, usually in lags 1 to 4. In the case of "I", about half of the markers that were uttered in close proximity consisted on false starts, such as the following examples:

(1) I actually, I was talking to my father
(2) I, I kind of like feel bad about it
(3) I mean, um, I, I don't know, I accused…

Approximately half of the "I" words that were uttered in close proximity were part of colloquially acceptable statements, as demonstrated in the following examples:

(1) I think I was probably still mad at him
(2) I knew I was going to college
(3) I actually applied and I ….

In the case of emotion words, the likelihood of finding two emotion words in close proximity is even smaller (which explains the lower trough in Figure 5 compared to Figure 1), especially considering the fact that coordinated affective adjectives are rare in spoken English. When positive affect words did cluster together, they served a function of stressing the affect by repetitions. For example (positive emotion words are underlined):

(1) And then at times he was very calm and loving
(2) She is very into being popular and, and being liked by the guys
(3) That was very sentimental and very um nice that I still treasure

The psychological factor may be related to the etiology and motivation for these speech patterns and their roots in the different acculturation processes of the genders. For example, while both men and women share presumably the same syntactic/stylistic rules and avoid uttering a succession of emotion words, the males are shown here to have both stronger as well as longer avoidance. Hence, men's speech has less likelihood of re-uttering an emotion word, and this decreased likelihood lasts almost approximately 4 words, while that of women lasts between one and two words. Thus, The sensitivity of the LCA technique allows for distinguishing between stylistic rules and psychologically-motivated individual preferences. The statistical factor influences the smoothness of the curve and hence the ability to discern smaller differences between the two subject populations. This factor involves the number of subjects sampled in the study, in combination with the frequency of the marker(s) under examination. In marker-level analyses, the frequency of the marker is determined by the prevalence of the marker in spoken language, while in thematic-level analysis the frequency is influenced by the size of the dictionary. Larger dictionary sizes will increase the frequency of the marker, but may decrease its accuracy by including words that do not belong consistently to the category at hand. For example, the dictionary used in this study was designed mainly for written samples (Pennebaker and King, 1999). Hence, words like "pretty", "kind" and "like", used in written samples to convey affective qualities, had to be excluded considering their overwhelming use in spoken US English as modifiers and fillers (Bradac et al., 1995; Jucker and Smith, 1998; Andersen, 2000).

The impact of the statistical factor on the shape of the correlation can be demonstrated by comparing the data on "I" to that of emotion words. The word "I" appears approximately 3 times more frequently than emotional words. The effect on the noise level of the LCA graphs can be demonstrated by comparing Figure 1 to Figure 5. Figure 5, which had the same number of subjects as Figure 1 but used a marker 3 times less frequent, shows more random variation of the correlation level, as can be seen by the squiggly nature of the graph lines. In addition, this sampling problem caused the uncertainty around the average level of correlation to be bigger, as evinced from the larger error bars compared to Figure 1. This increase in variability is particularly noticeable in the male population, in which the sample size is relatively small and consists only of 45 men. The fact that gender differences were still detected within these sampling limitations shows that these differences are larger than the variability within subjects, at least for the markers examined here.

References

Andersen, Gisle. 2000. The role of the pragmatic marker like in utterance interpreta-
        tion. In G. Andersen & T. Frethem (Eds.), *Pragmatic Markers and Pro-
        positional Attitude*. Philadelphia, USA: John Benjamins, 17–38.
Anderson, Kristin J., and Campbell Leaper. 1998. Emotion talk between same- and
        mixed-gender friends: Form and function. *Journal of Language and Social
        Psychology,* 17(4):419–448.
Bilous, Frances R., and Robert M. Krauss. 1988. Dominance and Accommodation in
        the Conversational Behaviours of Same- and Mixed-Gender Dyads. *Lan-
        guage and Communication,* 8(3–4):183–194.
Brown, Penelope, and Steven C. Levinson. 1987. *Politeness: Some Universals in
        Language Usage* (Vol. 4). New York, NY: Cambridge University Press.
Cameron, Deborah, Fiona McAlinden, and Kathy O'Leary. 1989. Lakoff in context:
        The social and linguistic functions of tag questions. In J. Coates & D. Cam-
        eron (Eds.), *Women in their Speech Communities: New Perspectives on
        Language and Sex*. London, UK: Longman 1989.
Caplan, Paula J., and Jeremy B. Caplan. 1999. *Thinking Critically about Research on
        Sex and Gender* (2nd ed.). New York: Addison-Wesley Longman.
Cappella, Joseph N. 1994. The management of conversational interaction in adults
        and infants. In M. L. Knapp & G. R. Miller (Eds.), *Handbook of Interper-
        sonal Communication*. Thousand Oak, California: Sage, 380–418.
Carli, Linda L. 1990. Gender, language and influence. *Journal of Personality and
        Social Psychology,* 59(5): 941–951.
Crawford, Mary E. 1995. *Talking Difference: On Gender and Language*. Thousand
        Oaks, CA: Sage.
Crosby, Faye, and Linda Nyquist. 1977. The female register: An empirical study of
        Lakoff's hypothesis. *Language in Society,* 6(3):313–322.
Fishman, Pamela. 1990. Conversational insecurity. In D. Cameron (Ed.), *The Femi-
        nist Critique of Language*. London, UK: Routledge, 1–26.
Fitzpatrick, Mary Anne, Anthony Mulac, and Kathryn Dindia. 1995. Gender-
        preferential language use in spouse and stranger interaction. *Journal of
        Language and Social Psychology,* 14(1-2):18–39.
Giles, Howard, Anthony Mulac, James J. Bradac, and Patricia Johnson. 1987. Speech
        accommodation theory: The first decade and beyond. *Communication
        Yearbook,* 10:13–48.
Gleser, Goldine. C., Louis A. Gottschalk, and John Watkins. 1959. The relationship
        of sex and intelligence to choice of words: A normative study of verbal be-
        havior. *Journal of Clinical Psychology,* 15:182–191.
Gottschalk, Louis A. 1969. *The Measurement of Psychological States through the
        Content Analysis of Verbal Behavior*. Berkeley, CA: University of Califor-
        nia Press.
Haas, Adelaide. 1979. Male and female spoken language differences: Stereotypes and
        evidence. *Psychological Bulletin,* 86(3):616–626.
Holmes, Janet. 1990. *Women, Men and Politeness*. New York, NY: Longman.

Jucker, Andreas. H., & Sara W. Smith. 1998. And People Just You Know Like 'Wow': Discourse Markers as Negotiating Strategies. In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers: Descriptions and Theory.* Amsterdam: John Benjamins, 171–201.

Kessler, Ronald C., Katherine A. McGonagle, Shanyang Zhao, Christopher B. Nelson, Paula J. Caplan, and Jeremy B. Caplan. 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results for the National Comorbidity Survey. *Archives of General Psychiatry,* 51:8–19.

Labov, William. 1966. *The Social Stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.

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

Labov, William, and Joshua Waletzky. 1997. Narrative Analysis: Oral versions of personal experience. *Journal of Narrative and Life History,* 7(1–4):3–38.

Lakoff, Robin T. 1975. *Language and Woman's Place.* New York, NY: Harper & Row.

Lakoff, Robin T. 1977. Women's language. *Language and Style,* 10(4):222–247.

Macaulay, Marcia. 2001. Tough talk: Indirectness and gender in requests for information. *Journal of Pragmatics,* 33:293–316.

Manning, Christopher D., and Hinrich Schütze. 1999. collocations, *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT Press, 151–189.

McLachlan, Angus. 1991. The effects of agreement, disagreement, gender and familiarity on patterns of dyadic interaction. *Journal of Language and Social Psychology,* 10(3):205–212.

McMillan, Julie R., A. Clifton, Diane McGrath, and Wanda S. Gale. 1977. Women's language: Uncertainty or interpersonal sensitivity and emotionality? *Sex Roles,* 3(6):545–559.

Mondorf, Britta. 2002. Gender Differences in English Syntax. *Journal of English Linguistics,* 30(2):158–180.

Mulac, Anthony, James J. Bradac, and Pamela Gibbons. 2001. Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research,* 27(1):121–152.

Mulac, Anthony, and Torborg L. Lundell. 1986. Linguistic contributors to the gender-linked language effect. *Journal of Language and Social Psychology,* 5(2):81–101.

Mulac, Anthony, and Torborg L. Lundell. 1994. Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language Effects. *Language and Communication,* 14(3):299–309.

Mulac, Anthony, Studley, L. B., and S. Blau. 1990. The gender-linked language effect in primary and secondary students' impromptu essays. *Sex Roles,* 23(9–10):439–469.

Mulac, Anthony, John M. Wiemann, Sally J. Widenmann, and Toni W. Gibson. 1988. Male/female language differences and effects in same-sex and

mixed-sex dyads: The gender-linked language effect. *Communication Monographs,* 55(4):315–335.

Oakes, Michael P. 1998. *Statistics for Corpus Linguistics*. Edinburgh, Scotland: Edinburgh University Press.

Pease, Barbara, and Allan Pease. 2001. *Why Men Don't Listen and Women Can't Read Maps: How We're Different and What to do About it*. Broadway Books.

Pennebaker, James W. 1995. Emotion, disclosure, and health: An overview. In: J. W. Pennebaker, Ed., *Emotion, Disclosure, and Health*. Washington, DC: American Psychological Association, 3–10.

Pennebaker, James W., and Martha E. Francis. 1999. Linguistic inquiry and word count: LIWC. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., and Laura A. King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology,* 77(6):1296–1312.

Ries, Klaus. 1999, September 5–9. Towards the detection and description of textual meaning indicators in spontaneous conversations. Paper presented at the European Conference on Speech Communication and Technology (EUROSPEECH '99), Budapest, Hungary.

Tannen, Deborah. 1991. *You Just Don't Understand: Women and Men in Conversation* (1st ed.). New York: Ballantine Books.

Tannen, Deborah. 1994a. *Gender and Discourse*. New York, NY: Oxford University Press.

Tannen, Deborah. 1994b. *Talking from 9 to 5: How Women's and Men's Conversational Styles Affect Who Gets Heard, Who Gets Credit, and What Gets Done at Work* (1st ed.). New York: W. Morrow.

Thomson, Robert, and Tamar Murachver. 2001. Predicting gender from electronic discourse. *British Journal of Social Psychology,* 40:193–208.

Thorne, Barrie, Cheris Kramarae, and Nancy Henley. 1983. Language, gender and society: Opening a second decade of research. In: B. Thorne, C. Kramarae, & N. Henley, Eds., *Language, Gender and Society*. Rowley, MA: Newbury House Publishers, Inc., 7–23.

Department of Psychology
New York University
6 Washington Place, Rm. #468
NYC, NY 10003
sc624@nyu.edu