Women Are More Likely Than Men to Use Tentative Language, Aren't They? A Meta-Analysis Testing for Gender Differences and Moderators

Campbell Leaper[1] and Rachael D. Robnett[1]

Abstract

Robin Lakoff proposed that women are more likely than men to use tentative speech forms (e.g., hedges, qualifiers/disclaimers, tag questions, intensifiers). Based on conflicting results from research testing Lakoff's claims, a meta-analysis of studies testing gender differences in tentative language was conducted. The sample included 29 studies with 39 independent samples and a combined total sample of 3,502 participants. Results revealed a statistically significant but small effect size ($d = .23$), indicating that women were somewhat more likely than men to use tentative speech. In addition, methodological moderators (operational definition, observation length, recording method, author gender, and year of study) and contextual moderators (gender composition, familiarity, student status, group size, conversational activity, and physical setting) were tested. Effect sizes were significantly larger in studies that (a) observed longer (vs. shorter) conversations, (b) sampled undergraduates (vs. other adults), (c) observed groups (vs. dyads), and (d) occurred in research labs (vs. other settings). The moderator effects are interpreted as supporting proposals that women's greater likelihood of tentative language reflects interpersonal sensitivity rather than a lack of assertiveness. In addition, the influence of self-presentation concerns in the enactment of gender-typed behavior is discussed.

The question of whether women and men differ in their language style has received much attention over the past several decades in psychology (e.g., Aries, 1996; Leaper & Ayres, 2007; Thorne & Henley, 1975), linguistics (e.g., Talbot, 1998; Tannen, 1994), communications (e.g., Dindia & Canary, 2006; Wood, 2007), and feminist studies (e.g., Crawford, 1995). Linguist Robin Lakoff (1973, 1975) played an important role in bringing this debate to the forefront. In a widely cited monograph originally published in *Language in Society*, Lakoff (1973) proposed that gender differences in communication are explained by the different roles that men and women hold in society. According to Lakoff, men communicate in an assertive manner because they occupy the dominant position in the social hierarchy. In contrast, she proposed women communicate in a more tentative and polite manner because they occupy the subordinate position in the social hierarchy. In the present study, we conducted a meta-analysis to test for average gender differences and possible moderators in the speech forms that Lakoff delineated. As reviewed later, the contextual moderators that we tested included gender composition, familiarity,

student status, group size, conversational activity, and physical setting. The methodological moderators that we considered were operational definition, observation length, recording method, author gender, and year of study. Before addressing the potential significance of these moderators, we begin with a review of Lakoff's (1973, 1975, 1977) proposals regarding women's and men's relative use of tentative language.

Comparing Women's and Men's Use of Tentative Language

A preliminary note on terminology is warranted: The phrases "women's language" (e.g., Lakoff, 1977; McMillan, Clifton, McGrath, & Gale, 1977) and "female register" (e.g., Crosby

---

[1] Department of Psychology, University of California, Santa Cruz, CA, USA

Corresponding Author:
Campbell Leaper, Department of Psychology, University of California, Santa Cruz, CA 95064, USA
Email: cam@ucsc.edu

& Nyquist, 1977) have been used to refer to the speech features that Lakoff (1973, 1975) highlighted in her earlier publications. This usage of these terms can inadvertently perpetuate an essentialist view whereby certain speech forms are characterized as inherently female. To avoid this practice, we shall refer to the investigated speech forms as "tentative language."

We considered four forms of tentative language that Lakoff had highlighted: expressions of uncertainty, hedges, tag questions, and intensifiers. First, *expressions of uncertainty* occur when speakers use disclaimers (e.g., "*I'm not sure if this is right*, but I think the meeting is tomorrow") or qualifiers (e.g., "Jim's performance in the course was *somewhat* disappointing"). Second, *hedges* include prefatory remarks such as *I guess* or modifiers such as *kind of* (e.g., "*I guess* the presentation was *kind of* short"). Lakoff suggested that women use hedges to downplay their authority. Third, *tag questions* are queries seeking confirmation of an immediately preceding declarative statement (e.g., "It's a beautiful day, *isn't it*?"). Lakoff proposed that women deploy tag questions to avoid being perceived as overly assertive when making a statement. Finally, *intensifiers* refer to adverbs such as *very, so*, or *really* used in a way that adds little content to a statement (e.g., "That report was *so* hard."). According to Lakoff, intensifiers mitigate the directness and strength of an assertion.

To test whether there are reliable gender differences in tentative language across different research reports, we conducted a meta-analysis. This is a technique for summarizing the statistical findings across studies investigating a similar outcome. A meta-analysis reveals whether there is a statistically significant effect across all studies. Another important component of the meta-analysis is the average effect size. Cohen's d is an index of effect size that reflects the difference between groups (e.g., women and men) in standard deviation units.

Besides testing for overall statistical significance and the average effect size, meta-analysis also allows researchers to test for the influence of moderator variables. For example, as explained in more depth later, we tested whether certain contextual and methodological factors influenced the likelihood and the size of any gender differences in tentative speech. Before addressing those points, we will provide an overview of the issues that warranted this meta-analysis.

Questions About Lakoff's Hypotheses

Lakoff's (1973, 1975, 1977) proposals regarding women's speech are widely cited in textbooks (e.g., Wood, 2007) and numerous empirical articles as evidence for gender differences in language. Decades later, many researchers credit Lakoff with sparking what has become a widely researched topic (see Lakoff & Bucholtz, 2004). Yet Lakoff's hypotheses have not been universally accepted. In the years following her initial publications, four major issues have been raised. The first limitation is that Lakoff's work was largely

speculative and therefore had little empirical basis (see Mulac & Bradac, 1995). Thus, her hypotheses may have reflected commonly held—and potentially inaccurate—stereotypes about women's speech patterns. Lakoff (1975, 1977) herself acknowledged this limitation, and she called upon researchers to empirically test her hypotheses. In a narrative review of the research literature, Aries (1996) concluded that there was mixed support for Lakoff's claims. She further suggested that gender differences depend on aspects of the interactive context and methodological features of particular studies. A meta-analysis can address these possibilities.

A second criticism frequently made about Lakoff's model is that it exaggerates and thereby essentializes gender differences in communication (Crawford, 1995; O'Barr & Atkins, 1980; Smith, 1985). More specifically, Lakoff has been challenged for suggesting that tentative language is used by the majority of women but is seldom used by men. A related criticism is that she overemphasized gender differences and thereby failed to acknowledge common similarities between men's and women's communication patterns. A meta-analysis may be especially helpful in this regard because effect sizes reflect the magnitude of difference and the corresponding degree of overlap between groups on a given measure.

Third, some psychologists have argued that tentative language may depend more on the relative status and power of the interaction partners than on their gender (Henley, 2001; LaFrance, 2001). From this perspective, tentative language is not necessarily characteristic of the feminine-stereotyped communication style. Rather, tentative language is used when someone—either male or female—is in a subordinate position. Supporters of this interpretation have proposed that "women's language" be renamed "powerless language" (e.g., O'Barr & Atkins, 1980). Thus, if women tend to use powerless language more often than do men, it may be because women are more likely than men to be in positions in which they lack power. In defense of Lakoff (1973, 1975), it is important to note that she attributed gender differences in speech to women's subordinate status. Indeed, her monograph was entitled, "Language and Woman's Place." Once again, meta-analysis may help to clarify this issue. As explained later in our section on moderators, male dominance may be implicated as an underlying factor if gender differences in tentative speech are more likely in mixed-gender than same-gender interactions.

Finally, some critics have worried that Lakoff's proposals imply that tentative language is somehow deficient; that is, tentative language might be viewed as substandard because it lacks assertiveness. A deficiency model plays into the greater social tendency to perceive feminine-stereotyped acts as problematic because these behaviors deviate from the masculine norm (i.e., "women-as-problem" perspective; Crawford, 1995). Along this vein, Spender (1984) noted that Lakoff used the masculine gender-typed speech pattern as the standard against which feminine gender-typed speech is

judged. According to Spender, women's and men's speech may indeed differ, but this difference does not necessarily mean that masculine gender-typed communication is superior. Instead, feminine gender-typed speech forms may have valuable functions. For instance, Fishman (1978) suggested that women use tag questions to keep others engaged in conversation. This understanding is consistent with an alternative proposal that women's language is more likely than men's language to reflect interpersonal sensitivity (e.g., McMillan et al., 1977). Therefore, it could be that women's use of tag questions and other forms of tentative speech reflect a greater emphasis on affiliation (Leaper & Ayres, 2007) rather than a lack of confidence.

Moderators of Gender Differences in Tentative Speech

In the years following Lakoff's (1973, 1975, 1977) initial publications, researchers sought to identify contextual factors that might elicit gender differences in communication style. Several of these factors were delineated in Leaper and Ayres' (2007) recent meta-analysis testing for gender differences and moderators in assertive and affiliative speech. Across studies, women were slightly less likely than men to use assertive speech ($d = .09$); women were also slightly more likely than men to use affiliative speech ($d = .12$). When the authors examined moderator variables, they discovered that the likelihood and the magnitude of gender differences in communication style depended on certain methodological characteristics and aspects of the social interaction.

Leaper and Ayres' (2007) meta-analysis synthesized much of the research pertaining to gender differences in communication. However, it did not specifically test for gender differences in tentative language as described by Lakoff (1973, 1975, 1977). The present meta-analysis is warranted because studies examining gender differences in communication have resulted in conflicting and complicated findings (e.g., Mulac, Lundell, & Bradac, 1986). The cumulative results of these studies are difficult to interpret because methodological and contextual factors unique to each study have yet to be examined as a whole. To this end, the present meta-analysis investigated a number of factors that may moderate gender differences in the use of tentative language. Methodological moderators include operational definition, length of observation, method of recording, gender of the first author, and year of study. In addition, we considered potential contextual moderators including gender composition, the relationship among conversational partners, student status, group size, conversational activity, and physical setting. These moderators are described more fully below.

Methodological moderators. Researchers' methodological choices may also affect the likelihood of observing gender differences in tentative language. We considered five possible methodological moderator variables. The first of these was *operational definition*. As reviewed earlier, Lakoff

(1973, 1975) proposed that women and men differ in their use of different forms of tentative speech. Although we tested for average gender differences across all forms of tentative language, it is possible that the likelihood and the size of differences vary across particular types of tentative speech (reflecting their operational definitions). Therefore, we compared the different operational definitions (qualifiers/disclaimers, hedges, tag questions, intensifiers) in our meta-analysis. The testing of this moderator was exploratory, and we did not advance any specific hypotheses.

Length of observation was a second factor that we considered. Research indicates that behavioral patterns become both easier to recognize and more consistent as the length of observation increases (e.g., Fagot, 1985). Hence, effect sizes may be greater when longer observation periods are made. *Method of recording* is a third factor that may moderate the findings of a study. Specifically, videotape can be a more accurate method of observation than is audiotape alone (Leaper & Ayres, 2007). If so, effect sizes may be stronger when videotape is employed. Fourth, we tested *gender of the first author* as a moderator. Past meta-analyses have demonstrated that this factor can influence the strength and direction of the effect size (e.g., Anderson & Leaper, 1998; Eagly & Carli, 1981). Therefore, we explored if this pattern occurred with regard to the set of studies that we sampled. Finally, we examined *year of study* to consider whether gender differences in the use of tentative language have changed over the years. In general, there is a recent historical trend toward smaller gender differences in assertiveness (Twenge, 2001), and we hypothesized a similar trend.

Contextual moderators. Six aspects of the conversational context were tested as possible moderators of gender differences in tentative language. First, the *gender composition* of the dyad or the group was of particular interest. If gender differences in tentative language reflect people's role expectations and men's greater dominance, then the magnitude of an average difference should be greater in mixed-gender than in same-gender interactions (see Carli, 1990; Hannover, 2000; Leaper & Ayres, 2007). Alternatively, gender differences in social behavior may reflect gender-typed social norms and preferences. If so, then differences between women and men should be greater when they are interacting with same-gender partners who are more apt to share similar norms (see Carli, 1990; Leaper & Ayres, 2007; McMillan et al., 1977; Palomares, 2009). To test these different possibilities, we compared studies involving same-gender interaction partners with studies involving other-gender interaction partners.

A second moderator that we tested was the *relationship among the conversational partners*. Past research indicates that people tend to behave in a more gender-typed manner when interacting with strangers than with familiar persons (Deaux & Major, 1987). This difference is likely related to the finding that many people anticipate more social approval

for gender-typed than cross-gender-typed behavior (see Deaux & Major, 1987; Leaper & Friedman, 2007). Hence, we compared studies that examined interactions between strangers to studies that examined interactions between familiar others. Larger average gender differences in tentative speech were expected among strangers than familiar partners.

Third, the *student status* of the participants was examined. The majority of psychological research is conducted with college-age participants, which raises the question of how well the results of these studies generalize to other age groups. Sampling is an important factor to consider when conducting research on gender norms (see Leaper & Ayres, 2007) because there is reason to believe that college students are less gender typed in their behavior than are other individuals. To this end, we compared studies conducted with undergraduate participants to studies conducted with other (usually older) participants. We hypothesized that average gender differences would be smaller among college than noncollege samples. Unfortunately, it was not possible to consider other sampling characteristics (e.g., ethnicity, culture, socioeconomic status), given the available studies.

A fourth contextual moderator that we tested was *group size*. Prior research suggests that dyadic interaction may foster intimacy, whereas larger groups may foster competitiveness (e.g., Benenson, Nicholson, Waite, Roy, & Simpson, 2001; Leaper & Ayres, 2007; Solano & Dunnam, 1985). Group size then has implications for participants' use of tentative language because past research has drawn parallels between tentative language and language that is used to establish intimacy (McMillan et al., 1977). If average gender differences in tentative speech reflect underlying differences in interpersonal sensitivity and intimacy, then effect sizes should be larger during dyadic than group interactions. Conversely, if variations in tentative speech reflect underlying gender differences in assertiveness and competitiveness, then the effect size should be larger during group than during dyadic interactions.

Fifth, we examined *conversational activity* as a moderator. The type of activity or conversational topic that women and men select to discuss may mediate some average gender differences in communication. On average, women are more likely than men to prefer personal topics and socioemotional activities, whereas men are more likely than women to prefer impersonal topics and task-oriented activities (e.g., Newman, Groom, Handelman, & Pennebaker, 2008). Gender differences may be mitigated, however, when women and men engage in similar activities or topics. Accordingly, we distinguished between structured activities (e.g., receiving an assigned topic to discuss) and unstructured activities (e.g., allowing participants to talk about whatever they want). We expected a stronger average gender difference during unstructured than structured conversations.

Finally, the *physical setting* was the sixth contextual moderator in our analyses. The majority of psychological research is carried out in controlled settings (e.g., university research laboratories); it is much less common for research to be conducted in naturalistic settings. Setting may be important because gender-typed behavior may be more likely in unfamiliar situations than in familiar situations (Deaux & Major, 1987). Therefore, we compared the results of studies carried out in labs to studies carried out in naturalistic settings. We expected that average gender differences would be more pronounced in lab settings.

Method

Literature Search

The majority of the studies used in our meta-analysis were found through the PsycINFO database by searching for studies that examined people's use of tentative language (defined below). In addition, we used other databases including Linguistics and Language Behavior Abstracts, Educational Resources Information Center (ERIC), and Sociological Abstracts. Some of the search terms included *tentative, assertive, powerless*, or *sensitivity* combined with terms such as *language, speech, communication*, or *social interaction*. The specific language forms that we examined (described later) were also used as search terms. Furthermore, we conducted additional searches of articles that cited either Robin Lakoff or key studies included in our meta-analysis. These citations as well as various reviews and books yielded other relevant studies.

The following criteria were used in determining whether a study would be included. First, only studies that examined both adult men's and adult women's use of tentative language were included; thus, studies were excluded if their samples comprised only one gender or children. Second, only studies that used observational methods to record participants' use of tentative language were included; thus, studies using self-report measures were not used. Third, we included only studies that examined face-to-face communication. Finally, only studies conducting statistical tests of gender differences were used; that is, qualitative studies were excluded. A total of 29 studies met these criteria for inclusion. Of these studies, publication dates ranged from 1977 to 2008. Information about each study is summarized in the appendix.

File-Drawer Problem

The "file-drawer problem" refers to the assumption that there is a bias toward publishing significant and/or compelling results, which means that null findings remain unpublished. Solely examining published studies, therefore, could paint a skewed picture of the overall effect of a phenomenon. Accordingly, some researchers advocate including unpublished studies in meta-analyses (e.g., Hedges & Vevea, 1996). We were able to identify two relevant unpublished dissertations that were included in the present meta-analysis. Furthermore, we utilized statistical techniques (described in results) to test for publication bias.

Coding the Studies

Trained research assistants worked individually to code the studies that met the criteria for inclusion. Specifically, the research assistants noted the statistical effects, operational definition, and other moderator variables (see below). The following statistical effects were recorded for each study: sample size, means and standard deviations (if available), and the test statistic (i.e., F, t, r, Z, p, χ^2). Coding for operational definition involved classifying the operational definitions used in each study into the five categories examined in the present analysis. Coding for the other moderator variables involved examining each study and noting whether the moderator variables under investigation in the present analysis were reported. If a moderator was present, the relevant values were coded. If a moderator was not present, this omission was noted, and the study was removed from the analysis of that particular moderator. The first author regularly met with the research assistants to discuss any questions that arose during the coding process. Disagreements were resolved through evaluation of the study in question and discussion. The second author double coded 25% of the studies. Inter-coder agreement was 100% for statistical effects, operational definition, and moderator variables.

Moderator Variables

Several broad classes of moderator variables were examined in the current meta-analysis. These included various methodological qualities as well as aspects of the interactive context. The specific moderators are described below.

Methodological moderators. Five methodological factors were tested as potential moderators of gender differences in tentative speech. First, this included the operational definition of tentative language. We compared different definitions based on Lakoff's (1975) model. Measures of tentative speech were classified as (a) *expressions of uncertainty* (qualifiers or disclaimers; e.g., "I'm not sure if this is right"), (b) *hedges* (e.g., "I guess," "kind of"), (c) *tag questions* (e.g., "It's hot today, isn't it?"), (d) *intensifiers* (adverbs such as "very," "so," or "really" when used in a way that adds little content to a statement), or (e) *general tentative language*. The latter category was included to incorporate studies that did not fall into any of the aforementioned categories but nonetheless examined tentative language. For example, some studies used a broader definition of tentative language that comprised more than a single category (e.g., a combination of qualifiers, hedges, and tag questions).

In addition to operational definition, four other methodological factors were examined: method for recording behavior (audiotape vs. videotape), length of observation, first author's gender, and publication year. For observation length, we contrasted interactions that were relatively short (1–10 min) and relatively long (11–75 min). For publication year, we used a median split to contrast studies from older (1977–1991) and

more recent (1992–2008) time periods. This allowed for approximately equal numbers of studies and roughly equal time spans within each period.

Contextual moderators. We examined the following six contextual factors: (a) *student status* (undergraduates vs. others), (b) the *relationship* between the participants (strangers vs. familiar), (c) the *gender composition* of the dyad or group (same- vs. mixed-gender), (d) the *size* of the group being observed (dyads vs. larger groups), (e) the observational *setting* (lab vs. other setting), and (f) the *conversational activity* or topic. Conversational activities were classified as either structured or unstructured. In structured activities, the participants were assigned specific activities or topics to discuss. Some examples included negotiation tasks, debates, and discussion of current events. In unstructured (or less structured activities), people were observed in situations where they selected their own conversation topics or activities.

Statistical Analyses

Effect sizes. The Comprehensive Meta-Analysis (CMA) statistical package was used to carry out the analyses. Cohen's *d*, which uses standard deviation units to measure the degree of difference between groups, was used as our index of effect size. For purposes of interpretation, Cohen (1988) suggested the following guidelines: Effect sizes are classified as "large" if $d \geq .8$ (reflecting 53% or less overlap between women and men), "medium" if *d* is between .5 and .8 (reflecting 66% or less overlap), or "small" if *d* is between .2 and .5 (reflecting less than 85% overlap). An effect size below .2 (reflecting more than 85% overlap) is considered negligible. In the current meta-analysis, average effects were positive if women were higher than men in the use of tentative language.

Random-effects versus mixed-effects models. The CMA software converts the inferential statistics used to test for a gender difference (e.g., *t, F, r, p*, or *M* and *SD*) into Cohen's *d* standardized effect measure. These standardized values are then combined using fixed-, random-, or mixed-effects models to create an overall effect size across studies (see below). On occasion, nonsignificant findings were not accompanied by sufficient information to allow for the computation of an effect size. (This was the case for about 15% of the effects.) When this shortcoming occurred, zero was imputed for the effect size. As described later in the results, we tested the overall effect size both with and without the studies where zero was imputed.

In the present analysis, we used a random-effects analysis to examine the overall effect across studies, and a mixed-effects analysis to examine moderator variables. Meta-analyses are often conducted with fixed-effects models, but these models are limited in the extent to which their results can be generalized to the population (that is, the body of hypothetical studies to which we would like to generalize our findings). In fixed-effects models, features of studies that may influence effect size are

assumed to be constant (i.e., fixed) across the population (Hedges, 1994). Therefore, all error variance is attributed to differences between the samples of participants in any given study. This is a limitation because it means that the findings from fixed-effects models should only be generalized to studies with identical predictor variables. Conversely, random-effects models are conducted under the assumption that the features of studies that influence effect size are randomly sampled from the population. Therefore, error variance is attributed both to the sampling of participants *and* to the sampling of predictor variables. This leads to enhanced generalizability because the results can be extended to studies beyond those that have identical predictor variables (Hedges, 1994). A mixed-effects model integrates features of the fixed- and random-effects models. Specifically, mixed-effects models are conducted under the assumptions that predictor variables (or moderators) are fixed across the population but that some random error variance remains after accounting for error due to sampling (Lipsey & Wilson, 2001).

CMA uses the Q_B statistic to test for the significance of moderator variables. For each condition associated with a moderator variable, an average effect size (d) and 95% confidence interval (CI), Z score, and Q_W statistic are reported. The Q_W statistic indicates whether there is significant heterogeneity of variance in effect sizes within a particular level of the moderator variable.

Trimming. A trimming procedure was used to examine the influence that outlier studies had on the pattern of results. Two separate analyses were performed to exclude the most extreme 10% and 20% of sampled studies.

Units of Analysis

The unit of analysis (k) in the meta-analysis was either the independent sample or the statistical test. As explained below, independent sample was used for all analyses except for one.

Independent sample as unit of analysis. Whenever a gender comparison was made for a specific group, it was treated as an independent sample. In most instances, a single research report counted as one independent sample. However, some studies reported effects separately for two or more conditions that constituted different samples. When this overlap occurred, the different groups were defined as independent samples and entered separately into the meta-analysis. The independent sample was the unit of analysis used to test for average differences across all studies.

Test as unit of analysis. In some independent samples, more than one type of tentative language was analyzed. For example, one study may have separately tested for gender differences in tag questions and hedges. Therefore, to examine operational definition as a moderator, we used the individual

statistical test as the unit of analysis. Consequently, independent samples with more than one operational definition were represented more than once when computing the average effect size in this analysis. When testing the other moderators, however, we averaged the effect sizes within an independent sample if more than one operational definition was tested.

Results

Test of Overall Gender Differences

There were 39 independent samples based on a combined total sample of 3,502 participants. A significant average effect size of small magnitude indicated that women were slightly more likely than men to use tentative language, $d = .23$, 95% CI = [.13, .32], $p < .001$. The test for homogeneity of variance was significant, $Q_W (38) = 62.04, p = .008$. Therefore, testing for moderator variables was warranted.

Trimming outliers. After trimming 10% of the outliers, the average effect size based on 35 independent samples ($N = 3,252$) was $d = .15$, 95% CI = [.08, .22], $p < .001$. After trimming 20% of outliers, the average effect size based on 31 independent samples ($N = 3,060$) was $d = .12$, 95% CI = [.05, .19], $p = .001$. Thus, trimming of 10% or 20% of the outliers reduced the magnitude of the average effect size.

Removing studies with zero-imputed effect size. As noted above, some studies with nonsignificant findings did not include sufficient information to allow for the computation of an effect size. This occurred for 6 (15%) of the studies included in our meta-analysis. In these situations, we imputed zero for the effect size. This conservative strategy may lead to an underestimation of the effect size. Hence, we also examined the overall effect without the studies that had imputed effect sizes. The average effect size based on 33 independent samples ($N = 2,970$) was $d = .27$, 95% CI = [.16, .37], $p < .001$. Thus, imputing a zero effect size for six nonsignificant studies only slightly underestimated the magnitude of the effect size (i.e., $d = .23$ versus $d = .27$). All of the available studies were therefore used when testing the moderator variables in subsequent analyses.

Tests for publication bias. We utilized funnel plot and trim-and-fill methods to test for publication bias. In the funnel plot (see Figure 1), the effect size for each study is plotted as a function of study size (measured by standard error). Visual inspection of the funnel plot indicates that the studies included in the present meta-analysis were more or less evenly distributed around the overall effect size; this pattern suggests that no publication bias was present (Egger, Davey, Schneider, & Minder, 1997). The trim-and-fill procedure (see Duval & Tweedie, 2000) builds on the information obtained in the funnel plot by predicting where missing studies are likely to fall. In this procedure, effect sizes for these missing studies are estimated; the overall effect size is then recomputed
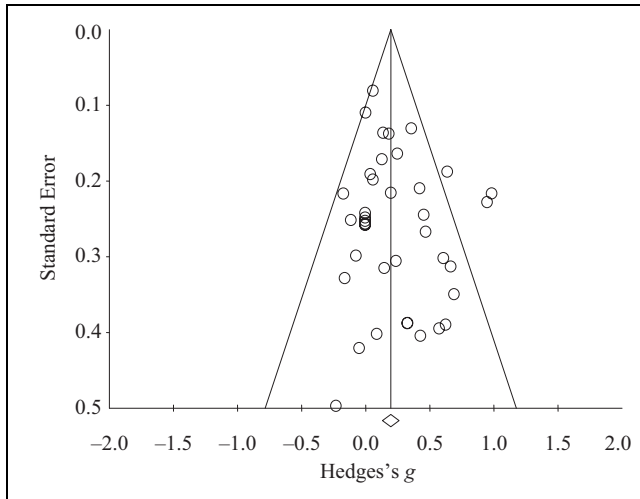
Figure 1. Effect size (Hedges's g) as a function of standard error. N = 39.

Table 1. Gender Effects on Tentative Speech by Operational Definition

Operational Definition	k	N	d	95% CI	Z	Q_W
General	16	982	.31	[.13, .49]	3.35**	25.64*
Uncertainty	14	1,075	.33	[.17, .50]	3.98***	22.20*
Hedges	12	1,799	.15	[.02, .29]	2.22*	15.61
Tag Questions	14	1,142	.23	[.07, .39]	2.86**	25.59*
Intensifiers	9	1,155	.38	[.14, .62]	3.12**	23.51**

Note. Operational definition was not a significant moderator of gender differences in tentative language, $Q_B = 4.25$, $p = .37$. Test was the unit of analysis, and k refers to the number of relevant tests for each operational definition. Positive effect sizes indicate a higher mean score for women than men. Q_W is a test for homogeneity of variance in effect sizes within a condition for a particular moderator.
* $p < .05$. ** $p < .01$. *** $p < .001$.

using these estimated effect sizes. The trim-and-fill procedure suggested that no studies were missing from the meta-analysis. Therefore, we concluded that there was no substantial evidence for publication bias in our meta-analysis.

Testing Moderators of Gender Differences

Methodological moderators. Operational definition was tested as a moderator with test as the unit of analysis. Analyses revealed that operational definition is not a significant moderator of gender differences in tentative language. Results for each operational definition are reported in Table 1.

Other tested methodological moderators were type of recording, observation length, first author gender, and publication year. The results for these moderators are summarized in Table 2. Observation length was a significant moderator of gender differences in tentative language: Gender differences were more evident in longer conversations ($d = .37$) than in shorter conversations ($d = .19$). There were no significant effects associated with first-author gender, publication year, or type of recording.

Contextual moderators. The results from the tests for the six contextual moderators are summarized in Table 2. There were no significant effects associated with participants' relationship (strangers vs. familiar), gender composition (same- vs. mixed-gender), or activity (structured vs. unstructured). However, student status and setting were both significant moderators; in addition, group size was a marginally significant ($p = .08$) moderator.

For student status, gender differences were more evident in studies of undergraduates ($d = .31$) than in studies of other populations ($d = .10$). For observational setting, gender differences were more evident for studies occurring in a research lab ($d = .27$) than for studies occurring in other locations ($d = .09$). For group size, gender differences were more

evident in groups ($d = .43$) than in dyads ($d = .18$); moreover, there was a significant correlation between group size and the magnitude of the effect, $r(37) = .40$, $p = .026$. Finally, although activity was not a significant moderator, the average effect size was significant for structured activities ($d = .25$) but not for unstructured activities ($d = .09$). We conducted follow-up tests to contrast specific types of structured activities with unstructured activities; however, we found no significant differences.

Associations among moderators. In order to examine whether any moderators were confounded with other moderators, we conducted a series of chi-square tests of independence. Chi-squares were used instead of correlations because our moderators were all categorical. Due to the large number of tests conducted, we set our α at .01 to reduce the likelihood of Type I error. Results indicated that there were two significant associations.

First, student status (undergraduate vs. other) and observational setting (lab vs. other) were significantly related, $\chi^2(1, N = 39) = 24.92$, $p < .001$. Among the studies sampling undergraduates ($k = 23$), *all* of them occurred in a research lab setting. Among the studies sampling non-undergraduates ($k = 16$), 75% of them occurred in a nonlab setting. Student status and observational setting were each previously indicated as significant moderators. Hence, one cannot disentangle the relative effect of each factor.

In addition, the variables of conversation partners (stranger vs. familiar) and length of observation (short vs. long) were significantly related, $\chi^2(1, N = 29) = 7.59$, $p = .006$. This test was conducted with studies that provided information about both observation length and partner familiarity. Among studies with a short observation length ($k = 18$), all of them were based on samples of strangers. Among studies with a longer observation length ($k = 11$), 64% were with strangers ($k = 7$) and the rest with familiar partners ($k = 4$). Among the studies with familiar partners that also indicated

Table 2. Methodological and Contextual Moderators of Gender Effects on Tentative Language

Moderator	k	N	d	95% CI	Z	Q_W	Q_B
Methodological factors							
Type of recording[a]							2.29
Audio	19	1887	.14	[.04, .24]	2.77**	20.71	
Video	17	1120	.30	[.13, .47]	3.38**	33.32**	
Observation length[b]							4.97*
1−10 min	18	1970	.19	[.08, .30]	3.30**	21.50	
11−75 min	14	760	.37	[.18, .57]	3.72***	26.80*	
First author gender							.17
Woman	21	2461	.24	[.11, .38]	3.45**	45.52**	
Man	18	1041	.21	[.10, .32]	3.66***	16.48	
Publication year							.19
1977−1991	20	1676	.21	[.07, .34]	2.97**	33.23*	
1992−2008	19	1826	.25	[.12, .37]	3.75***	28.57†	
Contextual factors							
Student status							6.14*
Undergraduates	23	1644	.31	[.17, .44]	4.40***	41.70**	
Other	16	1858	.10	[.01, .19]	2.17*	11.23	
Relationship[c]							.13
Strangers	30	2848	.23	[.12, .34]	4.24***	51.37**	
Familiar	5	174	.18	[−.10, .45]	1.26	4.16	
Gender composition[d]							.41
Same gender	5	209	.37	[−.12, .87]	1.47	12.19*	
Mixed gender	20	970	.21	[.10, .32]	3.82***	13.74	
Both	13	2233	.20	[.05, .36]	2.59*	30.73**	
Group size[e]							2.90†
Dyad	30	1940	.18	[.09, .26]	4.21***	30.88	
Group	8	1562	.43	[.15, .72]	2.99**	21.46**	
Observational setting[f]							5.70*
Lab	27	1552	.28	[.16, .41]	4.53***	47.19**	
Other	12	1562	.09	[−.01, .19]	1.80†	7.47	
Activity							2.67
Unstructured	11	651	.09	[−.06, .25]	1.17	7.86	
Structured	28	2851	.25	[.14, .36]	4.52***	52.06**	

Note. Independent sample was the unit of analysis, and *k* refers to the number of relevant independent samples for each condition. Positive effect sizes indicate a higher mean score for women than men. Q_B is an overall test of significance for a particular moderator. Q_W is a test for homogeneity of variance in effect sizes within a condition for a particular moderator.

[a] When testing type of recording as a moderator, three samples were excluded because they used on-site observations rather than either audio or video recordings.

[b] When testing length, seven samples were excluded because the length of the interaction was not specified.

[c] When testing relationship, four samples were excluded because either they used both familiar and unfamiliar participants or the familiarity was unclear.

[d] When testing gender composition, one sample was excluded because the gender composition was unclear.

[e] When testing group size, one sample was excluded because group size was not specified.

[f] Student status and observational setting were confounded moderators inasmuch as all studies observing undergraduate students occurred in a research lab setting.

† $p = .08$. * $p < .05$. ** $p < .01$. *** $p < .001$.

observation length ($k = 4$), all of them had a long observation length. Thus, familiarity and observation length were somewhat confounded. However, only length (and not familiarity) was a significant moderator of gender differences in tentative speech.

Discussion

The meta-analysis provided support for Lakoff's (1973, 1975, 1977) proposal that women are more likely than men

to use tentative speech forms. We considered general definitions for four types of tentative speech: hedges, expressions of uncertainty, intensifiers, and tag questions. In addition, we also included a fifth category of studies using a composite measure of tentative speech. The magnitude of the average difference across all measures ($d = .23$) is what Cohen (1988) considered small but meaningful. However, when outliers were removed, the average difference fell into the negligible range ($d = .15$ with 10% outliers removed).

Operational definition used by researchers to describe language was not a significant moderator of gender differences. Although operational definition did not significantly moderate effect sizes, there was a range in the magnitude of effect sizes among the specific types of tentative speech ($d = .15$ for hedges, $d = .23$ for tag questions, $d = .33$ for uncertainty, and $d = .38$ for intensifiers).

One way to consider the magnitude of difference is in the amount of overlap and nonoverlap between women's and men's distributions (Cohen, 1988). The significant effect sizes that we observed ranged from approximately .2 to .4. An effect size of $d = .2$ reflects 85% overlap (15% nonoverlap). An effect size of $d = .4$ reflects 73% overlap (27% nonoverlap). Although these effect sizes indicate meaningful differences, it is worth underscoring the great deal of overlap in the two genders' distributions. Recognizing the degree of overlap is important for two reasons: First, average gender differences in communication style are often exaggerated; and, second, highlighting any average gender difference can perpetuate an essentialist view of women and men as fundamentally different (see Crawford, 1995; O'Barr & Atkins, 1980; Smith, 1985). For the vast majority of women and men in the sampled studies, there was much more overlap than difference in the use of tentative speech. Furthermore, as Hyde (2005) highlighted in her review of various meta-analyses testing for gender differences, such overlap is common for most social behaviors.

Besides operational definition, we tested first-author gender, publication year, type of recording, and observation length as potential methodological moderators of gender differences in tentative speech. Author gender was tested as a check for possible gender bias (e.g., see Eagly & Carli, 1981). None was indicated. With regard to publication year, we hypothesized that average gender differences in tentative speech would be less likely in more recent studies. This prediction was based on Twenge's (2001) meta-analysis, which indicated that gender differences in self-reported assertiveness had declined during recent decades. However, contrary to expectations, this moderator variable was not significant in our meta-analysis. As discussed later, tentative language may reflect interpersonal sensitivity more than lack of assertiveness; if so, then cultural changes in women's assertiveness may be less apparent with regard to these speech forms.

Length of observation was a significant moderator. Larger differences were associated with longer observation periods (11 or more min). This finding suggests that very brief periods of observation may not be sufficiently sensitive to detect certain aspects of people's communication style. We hypothesized that type of recording might have an analogous effect with video recording being more sensitive (i.e., more likely to detect significant effects) than audio recording. However, the two methods were not significantly different, which may mean that the recording method is less relevant when examining speech behavior.

The contextual moderators offer a more interesting way of understanding observed gender differences in tentative speech. One potentially revealing moderator is the gender composition of the interaction partners. According to Lakoff's (1973, 1975, 1977) original proposal, gender differences in tentative speech reflect women's subordinate status relative to men; that is, tentative speech is used to downplay power in a social interaction. If this interpretation was correct, we would expect a larger gender difference during mixed-gender than same-gender interactions. For example, Leaper and Ayres (2007) found that gender differences in talkativeness were larger in mixed-gender than same-gender interactions; they suggested this difference lends support to previous proposals that some men use speech to dominate women in conversation. In the present meta-analysis, however, we did not find support for this explanation of gender differences in tentative speech. If anything, the magnitude of gender difference was slightly (but not significantly) larger during same-gender ($d = .37$) than mixed-gender ($d = .21$) interactions. This pattern lends support to an alternative interpretation regarding the meaning of tentative speech.

Rather than reflecting lower power and status, tentative speech can function to express interpersonal sensitivity (McMillan et al., 1977). When a speaker softens an assertion through the use of a qualifier or a tag question, she or he is seeking the listener's consent and involvement. For example, a tag question explicitly invites the listener to respond. Interpersonal sensitivity is generally emphasized more during girls' than boys' gender socialization (Leaper & Friedman, 2007). The social norm emphasizing interpersonal sensitivity among girls and women is not mutually exclusive with institutionalized male dominance. To the extent that women and girls hold a more subordinate status in society, they learn it is important to be sensitive to interpersonal cues (Henley, 2001). These alternative functions for tentative speech may explain why we did not see a significant difference between same-gender and mixed-gender interactions. That is, gender-typed social norms for interpersonal sensitivity may be salient in women's same-gender interactions, whereas male dominance may occur in cross-gender interactions. Thus, there may be different reasons for average gender differences in some social behaviors operating in same-gender and mixed-gender interactions. We encourage researchers to explore these speculations in future studies.

Group size was a marginally significant moderator of gender differences in tentative speech. There was a negligible effect size during dyadic interactions ($d = .18$), but a small-to-moderate effect size during group interactions ($d = .43$). In general, individuals are more likely to treat one another as equals in dyadic than group interactions; therefore, gender differences in interpersonal sensitivity may be reduced in dyadic contexts. Group interactions, however, are more likely to elicit competition for people's attention and viewpoints (Leaper & Ayres, 2007). Accordingly, gender differences in interpersonal sensitivity may be more likely to emerge in more competitive settings.

When activity was taken into account, there was not a significant difference between structured ($d = .25$) and unstructured ($d = .09$) activities—although the average effect was significant during the former but not the latter condition. This finding contradicts the contextual model that the type of activity mediates gender differences in social behavior. For example, according to this model, women may use more affiliative speech forms because they are more likely to discuss personal matters, whereas men may be more likely to use more assertive speech forms because they engage in more task-oriented activities (Hall & Mast, 2008). Rather than seeing significantly larger effect sizes in the unstructured than structured condition, the trend was in the other direction: Effect sizes were slightly larger in the structured context (see Leaper & Ayres, 2007, for a similar pattern regarding average gender differences in assertive speech). The structured activities in most of the sampled studies comprised assigned instrumental tasks such as debating issues, negotiating plans, or solving a puzzle. These are relatively masculine gender-typed contexts that may have made gender roles more salient for the participants. Thus, some men's desire to establish their authority may have superseded their concerns with interpersonal sensitivity. Conversely, some women may have interpreted these activities as requiring cooperation and interpersonal sensitivity. This supposition requires testing in future research (see Palomares, 2009). One approach is to consider if and how interpersonal goals mediate gender differences in communication style (see Burleson, 2002; Palomares, 2009; Strough & Berg, 2000).

Self-presentation is one kind of interpersonal goal that may underlie gender-related variations in the use of tentative speech. In this regard, researchers find that self-presentational concerns tend to be heightened in unfamiliar situations. When this occurs, people sometimes rely on gender-role stereotypes to guide their behavior (Deaux & Major, 1987). Contrary to the self-presentation model, we did not uncover a difference between studies of strangers and familiar persons. This absence of an effect may have been due to the small number of samples ($k = 5$) looking at familiar partners. Other contextual moderators, however, indicated support for the possible influence of self-presentation. These included group size and observational setting. Larger gender differences in tentative speech occurred during group ($d = .43$) than during dyadic ($d = .18$) interactions as well as in research labs ($d = .28$) than in other settings ($d = .09$). As Deaux and Major (1987) reviewed, greater uncertainty and corresponding concerns with self-presentation tend to occur in larger groups. Also, similar concerns are more likely in unfamiliar settings, and a research laboratory is an unusual context for a conversation.

We also detected larger differences among studies sampling undergraduate students ($d = .31$) than other populations ($d = .10$). On one hand, this result is surprising because one might expect college students to be relatively egalitarian in their gender roles and behavior. On the other hand, young adults may be especially concerned with self-presentational concerns because they are exploring their identities and looking to others for validation. The latter pattern may decline during the course of adulthood (Eaton, Mitchell, & Jolley, 1991), which may partly explain why average effect sizes were smaller in older noncollege samples. Another point to consider is that undergraduate samples were exclusively observed in research lab settings. Hence, the unfamiliar lab setting may have further accentuated—or possibly accounted for—the greater likelihood of gender differences in tentative speech among undergraduates than other nonstudents.

Turning to the limitations of our study, the kinds of analyses that we could conduct were constrained by the number of available studies. First, the moderators that we could test were limited to those included in the various studies. Some potentially interesting moderators such as ethnicity, culture, or socioeconomic status could not be examined because these factors have not been tested and did not vary sufficiently across studies. A related limitation was that we were not able to differentiate more specifically among different levels or conditions of our moderator variables. For example, our analysis of relationship type was limited to strangers versus familiar persons. If there were several more relevant studies, it might be revealing to compare friends, dating partners, spouses, and coworkers. Also, as discussed earlier, the types of assigned activities in the studies were mostly task-oriented and did not include self-disclosure or other more feminine-stereotyped situations (see Leaper & Ayres, 2007, for possible examples regarding other speech behaviors). Finally, we were not able to consider possible interaction effects among multiple moderator variables.

Despite the aforementioned limitations, we believe our meta-analysis makes a useful contribution to our understanding of gender-related variations in language. We built on prior narrative reviews of the research literature addressing Lakoff's (1975, 1977) proposals regarding so-called women's language (e.g., Aries, 1996; Crawford, 1995; LaFrance, 2001). Through our use of quantitative meta-analysis, we found support for Lakoff's hypothesis that women are more likely than men to use tentative speech. The moderator effects are viewed as compatible with the proposal that tentative speech reflects interpersonal sensitivity rather than a lack of assertiveness. In addition, we highlighted the importance of self-presentational concerns.

The average gender difference in tentative language was statistically significant, although the magnitude of the effect was small. Small effects can have important consequences when they are sustained over long periods of time (see Abelson, 1985; Eagly, 1995). There is evidence suggesting that the use of some tentative speech forms may be stable for speakers (Bradac, Mulac, & Thompson, 1995). If some women use tentative speech at slightly higher rates than most men, then this may make a difference in how these women affect their listeners. Relative to those who almost never use tentative speech, these women may be viewed as being either

polite or unassertive. Furthermore, perceptions of tentative speech as reflecting either interpersonal sensitivity or power-lessness may vary according to the listener's gender—with women more likely to interpret tentative speech as a sign of interpersonal sensitivity and men more likely to view it as a lack of assertiveness (e.g., Mulac et al., 1998). These differing interpretations may contribute to miscommunication, power asymmetries, and relationship dissatisfaction in cross-gender interactions. The effect may be compounded when there are other average gender differences in communication style (e.g., Leaper & Ayres, 2007).

Whereas a small average gender difference in tentative language may be meaningful in some social interactions and relationships, it is important to reiterate that there was not a pervasive gender difference. A great deal of overlap occurred between women and men in their uses of tentative speech. This means that many men used tentative speech with equal or even greater frequency than did the average woman. Furthermore, the likelihood of a significant gender difference in tentative language was context-dependent. As the moderator analyses revealed, there was a small gender difference in some situations, whereas there was a negligible difference in other contexts. Our meta-analysis supports the view that women and men are more similar than different (Hyde, 2005). Accordingly, tentative speech should be viewed as both women's *and* men's language.

Authors' Note

Declaration of Conflicting Interests

Funding

References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133.

Anderson, K. J., & Leaper, C. (1998). Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. *Sex Roles*, *39*, 225–252.

Aries, E. (1996). *Men and women in interaction: Reconsidering the difference.* New York, NY: Oxford University Press.

Benenson, J. F., Nicholson, C., Waite, A., Roy, R., & Simpson, A. (2001). The influence of group size on children's competitive behavior. *Child Development*, *72*, 921–928.

*Bradac, J., Mulac, A., & Thompson, S. A. (1995). Men's and women's use of intensifiers and hedges in problem-solving interaction: Molar and molecular analyses. *Research on Language and Social Interaction*, *28*, 93–116.

*Brouwer, D., Gerritsen, M., & DeHaan, D. (1979). Speech differences between women and men: On the wrong track? *Language in Society*, *8*, 33–50.

Burleson, B. R. (2002). Psychological mediators of sex differences in emotional support: A reflection on the mosaic. *Communication Reports*, *15*, 71–79.

*Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology*, *59*, 941–951.

*Case, S. S. (1988). Cultural differences, not deficiencies: An analysis of managerial women's language. In S. Rose & L. Larwood (Eds.), *Women's careers: Pathways and pitfalls* (pp. 41–63). New York, NY: Praeger.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crawford, M. E. (1995). *Talking difference: On gender and language.* Thousand Oaks, CA: Sage.

*Crosby, F., Jose, P., & Wong-McCarthy, W. (1981). Gender, androgyny, and conversational assertiveness. In C. Mayo & N. M. Henley (Eds.), *Gender and nonverbal behavior* (pp. 151–169). New York, NY: Springer-Verlag.

*Crosby, F., & Nyquist, L. (1977). The female register: An empirical study of Lakoff's hypotheses. *Language in Society*, *6*, 313–322.

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review*, *94*, 369–389.

Dindia, K., & Canary, D. J. (2006). *Sex differences and similarities in communication* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

*Dixon, J. A., & Foster, D. H. (1997). Gender and hedging: From sex differences to situated practice. *Journal of Psycholinguistic Research*, *26*, 89–107.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, *50*, 145–158.

Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influence-ability: A meta-analysis of social influence studies. *Psychological Bulletin*, *90*, 1–20.

Eaton, Y. M., Mitchell, M. L., & Jolley, J. M. (1991). Gender differences in the development of relationships during late adolescence. *Adolescence*, *26*, 565–568.

Egger, M., Davey, S. G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.

Fagot, B. I. (1985). A cautionary note: Parents' socialization of boys and girls. *Sex Roles*, *12*, 471–476.

Fishman, P. M. (1978). Interaction: The work women do. *Social Problems*, *25*, 397–406.

*Grob, L. M., Meyers, R. A., & Schuh, R. (1997). Powerful/ powerless language use in group interactions: Sex differences or similarities? *Communication Quarterly*, *45*, 282–303.

Hall, J. A., & Mast, M. S. (2008). Are women always more interpersonally sensitive than men? Impact of goal and content domain. *Personality and Social Psychology Bulletin*, *34*, 144–155.

*Hannah, A., & Murachver, T. (2007). Gender preferential responses to speech. *Journal of Language and Social Psychology*, *26*, 274–290.

Hannover, B. (2000). Development of the self in gendered contexts. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 177–206). Mahwah, NJ: Lawrence Erlbaum.

Hedges, L. V. (1994). Statistical considerations. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29–38). New York, NY: Russell Sage Foundation.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, *21*, 299–332.

Henley, N. M. (2001). Body politics. In A. Branaman (Ed.), *Self and society* (pp. 288–297). Malden, MA: Blackwell Publishers.

*Hladik, E. G., & Edwards, H. T. (1984). A comparative analysis of mother-father speech in the naturalistic home environment. *Journal of Psycholinguistic Research*, *13*, 321–332.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592.

*Johnson, C. (1994). Gender, legitimate authority, and leader-subordinate conversations. *American Sociological Review*, *59*, 122–135.

*Johnson, C., Funk, S. J., & Clay-Warner, J. (1998). Organizational contexts and conversation patterns. *Social Psychology Quarterly*, *6*, 361–371.

*Kollock, P., Blumstein, P., & Schwartz, P. (1985). Sex and power in interaction: Conversational privileges and duties. *American Sociological Review*, *50*, 34–46.

LaFrance, M. (2001). Gender and social interaction. In R. K. Unger (Ed.), *Handbook of the psychology of women and gender* (pp. 245–255). Hoboken, NJ: Wiley.

Lakoff, R. (1973). Language and woman's place. *Language in Society*, *2*, 45–79.

Lakoff, R. (1975). *Language and woman's place.* New York, NY: Harper & Row.

Lakoff, R. (1977). Women's language. *Language and Style*, *10*, 222–247.

Lakoff, R., & Bucholtz, M. (Ed.). (2004). *Language and woman's place: Text and commentaries.* New York, NY: Oxford University Press.

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review*, *11*, 328–363.

Leaper, C., & Friedman, C. K. (2007). The socialization of gender. In J. Grusec & P. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 561–587). New York, NY: Guilford.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis: Applied social research methods series* (Vol. 49). Thousand Oaks, CA: Sage.

*Makay, D. L. (1993). *The impact of sex and gender on powerful/ powerless language use and perceptions of credibility during an ongoing dyadic conversation* (Unpublished doctoral dissertation). Ohio State University, Columbus, Ohio, USA.

*Martin, J. N., & Craig, R. T. (1983). Selecting linguistic sex differences during initial social interactions of same-sex and mixed-sex student dyads. *The Western Journal of Speech Communication*, *47*, 16–28.

*McFayden, R. G. (1996). Gender, status, and "powerless" speech: Interaction of students and lecturers. *British Journal of Social Psychology*, *35*, 353–367.

*McMillan, J. R., Clifton, A. K., McGrath, D., & Gale, W. S. (1977). Women's language: Uncertainty or interpersonal sensitivity and emotionality? *Sex Roles*, *3*, 545–559.

*McMullen, L. M., Vernon, A. E., & Murton, T. (1995). Divisions of labor in conversations: Are Fishman's results replicable and generalizable? *Journal of Psycholinguistic Research*, *24*, 255–268.

*Moore, S. F., Shaffer, L., Goodsell, D. A., & Baringoldz, G. (1983). Gender or situationally determined spoken language differences? The case of leadership situation. *International Journal of Women's Studies*, *6*, 44–53.

Mulac, A., & Bradac, J. J. (1995). Women's style in problem solving interactions: Powerless, or simply feminine? In P. J. Kalbfleish & M. J. Cody (Eds.), *Gender, power and communication* (pp. 83–104). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mulac, A., Erlandson, K. T., Farrar, W. J., Hallett, J. S., Molloy, J. L., & Prescott, M. E. (1998). "Uh-huh. What's that all about?" Differing interpretations of conversational backchannels and questions as sources of miscommunication across gender boundaries. *Communication Research*, *25*, 641–668.

Mulac, A., Lundell, T. L., & Bradac, J. J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs*, *53*, 115–129.

*Mulac, A., Seibold, D. R., & Farris, J. L. (2000). Female and male managers' and professionals' criticism and giving: Differences in language use and effects. *Journal of Language and Social Psychology*, *19*, 389–415.

*Mulac, A., Wiemann, J. M., Widenmann, S. J., & Gibson, T. W. (1988). Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communication Monographs*, *55*, 315–335.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*, 211–236.

O'Barr, W., & Atkins, B. (1980). "Women's language" or "powerless language"? In S. McConnell-Ginet, R. Borker, & N. Furman (Eds.), *Women and language in literature and society* (pp. 93–110). New York, NY: Praeger.

Palomares, N. A. (2009). Women are sort of more tentative than men, aren't they? How men and women use tentative language differently, similarly, and counterstereotypically as a function of gender salience. *Communication Research*, *36*, 538–560.

*Precht, K. (2008). Sex similarities and differences in stance in informal American conversation. *Journal of Sociolinguistics*, *12*, 89–111.

*Reid, S. A., Keerie, N., & Palomares, N. A. (2003). Language, gender salience and social influence. *Journal of Language and Social Psychology*, *22*, 210–233.

*Rubin, D. L., & Nelson, M. W. (1983). Multiple determinants of stigmatized speech style: Women's language, powerless language, or everyone's language. *Language and Speech*, *26*, 273–290.

*Sayers, F., & Sherblom, J. (1987). Qualifications in male language as influenced by age and gender of conversation partner. *Communication Reports*, *4*, 88–92.

*Smeltzer, L. R., & Watson, K. W. (1986). Gender differences in verbal communication during negotiation. *Communication Research Reports*, *3*, 74–79.

Smith, P. M. (1985). *Language, the sexes and society.* New York, NY: Basil Blackwell.

Solano, C. H., & Dunnam, M. (1985). Two's company: Self-disclosure and reciprocity in triads versus dyads. *Social Psychology Quarterly*, *48*, 183–187.

Spender, D. (1984). *Man-made language.* London, UK: Routledge & Kegan Paul.

*Stiles, W. B., Lyall, L. M., Knight, D. P., Ickes, W., Waung, M., Hall, C. L., . . . Primeau, B. E. (1997). Gender differences in verbal presumptuousness and attentiveness. *Personality and Social Psychology Bulletin*, *7*, 759–772.

Strough, J., & Berg, C. A. (2000). Goals as a mediator of gender differences in high-affiliation dyadic conversations. *Developmental Psychology*, *36*, 117–125.

Talbot, M. M. (1998). *Language and gender.* Malden, MA: Blackwell Publishers.

Tannen, D. (1994). *Gender and discourse.* New York, NY: Oxford University Press.

Thorne, B., & Henley, N. (1975). *Language and sex: Difference and dominance.* Rowley, MA: Newbury House.

Twenge, J. M. (2001). Changes in women's assertiveness in response to status and roles: A cross-temporal meta-analysis, 1931-1993. *Journal of Personality and Social Psychology*, *81*, 133–145.

*Whitney, D. L. (1991). "*Women's" or "powerlesss" language in conversation: Gender, status, and attributional consequences* (Unpublished doctoral dissertation). University of Saskatchewan, Saskatoon, Canada.

Wood, J. T. (2007). *Gendered lives: Communication, gender, and culture* (7th ed.). Belmont, CA: Wadsworth.

References marked with an asterisk indicate studies included in the meta-analysis.

Appendix

Study Characteristics for Tentative Language by Statistical Test

Author and Year	N (w)	N (m)	Stat Value	d	Op Def	Author Gender	Obs Length	Record	Sample	Relation	Gender Comp	Size	Setting	Activity
Bradac et al. (1995) a	58	58	$t = 3.27$	.61	IN	M	10	A	UG	Str	Both	2	Lab	Strx
Bradac et al. (1995) b	58	58	$t = 3.58$	.67	HE	M	10	A	UG	Str	Both	2	Lab	Strx
Brouwer, Gerritsen, and DeHaan (1979)	309	278	$p = .049$	.07	UN	W	3	A	Oth	Str	Both	2	Oth	Strx
Carli (1990) a1	30	30	$p = 1$	0	GEN	W	10	V	UG	Str	SG	2	Lab	Strx
Carli (1990) a2	30	30	$p = 1$	0	UN	W	10	V	UG	Str	SG	2	Lab	Strx
Carli (1990) a3	30	30	$p = 1$	0	HE	W	10	V	UG	Str	SG	2	Lab	Strx
Carli (1990) a4	30	30	$p = 1$	0	TQ	W	10	V	UG	Str	SG	2	Lab	Strx
Carli (1990) a5	30	30	$t = 2.21$	.57	IN	W	10	V	UG	Str	SG	2	Lab	Strx
Carli (1990) b1	58	58	$t = 4.42$	.58	GEN	W	10	V	UG	Str	MG	2	Lab	Strx
Carli (1990) b2	58	58	$t = 4.36$	.57	UN	W	10	V	UG	Str	MG	2	Lab	Strx
Carli (1990) b3	58	58	$t = 2.83$	.37	TQ	W	10	V	UG	Str	MG	2	Lab	Strx
Carli (1990) b4	58	58	$t = 2.83$	.31	HE	W	10	V	UG	Str	MG	2	Lab	Strx
Carli (1990) b5	58	58	$p = 1$	0	IN	W	10	V	UG	Str	MG	2	Lab	Strx
Case (1988) a	5	5	$p = .09$	.63	UN	W	45	V	Oth	Oth	MG	10	Oth	Strx
Case (1988) b	5	5	$p = .09$	.63	TQ	W	45	V	Oth	Oth	MG	10	Oth	Strx
Crosby and Nyquist (1977) (Study 1)	16	16	$t = 1.96$	.69	GEN	W	12	A	UG	Oth	SG	2	Lab	Strx
Crosby and Nyquist (1977) (Study 2)	90	107	$p = 1$	0	GEN	W	1	O	Oth	Str	MG	2	Oth	Strx
Crosby and Nyquist (1977) (Study 3)	309	278	$F = .41$	.43	GEN	W	2	O	Oth	Str	Oth	2	Oth	Strx
Crosby, Jose, and Wong-McCarthy (1981)	48	48	$p = 1$	0	GEN	W	6	A	UG	Str	MG	2	Lab	Strx
Dixon and Foster (1997)	54	50	$F = .48$	−.14	HE	M	6	V	UG	Str	Both	2	Lab	Strx
Grob, Meyers, and Schuh (1997) a	44	20	$\chi^2 = 4.73$	.61	UN	W	n/a	V	UG	Str	MG	5	Lab	Strx

(continued)

Appendix (continued)

Author and Year	N (w)	N (m)	Stat Value	d	Op Def	Author Gender	Obs Length	Record	Sample	Relation	Gender Comp	Size	Setting	Activity
Grob et al. (1997) b	44	20	$\chi^2 = 3.58$	.53	HE	W	n/a	V	UG	Str	MG	5	Lab	Strx
Grob et al. (1997) c	44	20	$\chi^2 = 1.08$	.28	TQ	W	n/a	V	UG	Str	MG	5	Lab	Strx
Hannah and Murachver (2007)	24	24	$p = 1$	0	TQ	F	8	A	Oth	Str	MG	2	Oth	Unst
Hladik and Edwards (1984)	10	10	M & SD	−.05	TQ	W	30	A	Oth	Oth	MG	3	Oth	Strx
Johnson (1994)	71	70	$p = .05$	0	UN	W	10	V	UG	Str	Both	3	Lab	Strx
Johnson, Funk, and Clay-Warner (1998)	40	40	M & SD	.95	GEN	W	11	V	UG	Str	SG	4	Lab	Strx
Kollock, Blumstein, and Schwartz (1985)	15	15	$p = 1$	0	TQ	M	n/a	A	Oth	Oth	MG	2	Oth	Strx
Makay (1993) a	66	66	$F = .28$	−.02	TQ	M	10	V	UG	Str	MG	2	Lab	Strx
Makay (1993) b	66	66	$F = 2.58$	.43	GEN	M	10	V	UG	Str	MG	2	Lab	Strx
Martin and Craig (1983)	20	20	$F = 4.46$	.67	UN	W	4	V	UG	Str	Both	2	Lab	Unst
McFayden (1996) a	103	105	M & SD	.24	HE	W	15	A	Oth	Str	Both	2	Lab	Strx
McFayden (1996) b	103	105	M & SD	0	UN	W	15	A	Oth	Str	Both	2	Lab	Strx
McFayden (1996) c	103	105	$F = 6.0$	.34	TQ	W	15	A	Oth	Str	Both	2	Lab	Strx
McMillan, Clifton, McGrath, and Gale (1977) a	61	37	$t = 3.98$	.99	UN	W	30	V	UG	Str	Both	6	Lab	Strx
McMillan et al. (1977) b	61	37	$t = 4.94$	1.03	TQ	W	30	V	UG	Str	Both	6	Lab	Strx
McMillan et al. (1977) c	61	37	$t = 5.29$	1.10	IN	W	30	V	UG	Str	Both	6	Lab	Strx
McMullen, Vernon, and Murton (1995)	17	17	$t = −.126$	−.14	TQ	W	30	V	Oth	Oth	MG	2	Oth	Unst
Moore, Shaffer, Goodsell, and Baringoldz (1983) a	20	20	$F = 2.56$	−.38	UN	M	n/a	A	UG	Str	Both	2	Lab	Strx
Moore et al. (1983) b	20	20	$p = 1$	0	TQ	M	n/a	A	UG	Str	Both	2	Lab	Strx
Mulac, Wiemann, Widenmann, and Gibson (1988) a1	12	12	M & SD	.43	UN	M	20	A	UG	Str	SG	2	Lab	Strx
Mulac et al. (1988) a2	12	12	M & SD	.29	HE	M	20	A	UG	Str	SG	2	Lab	Strx
Mulac et al. (1988) a3	12	12	M & SD	.29	IN	M	20	A	UG	Str	SG	2	Lab	Strx
Mulac et al. (1988) b1	24	24	M & SD	.05	UN	M	20	A	UG	Str	MG	2	Lab	Strx
Mulac et al. (1988) b2	24	24	M & SD	−.34	HE	M	20	A	UG	Str	MG	2	Lab	Strx
Mulac et al. (1988) b3	24	24	M & SD	.48	IN	M	20	A	UG	Str	MG	2	Lab	Strx
Mulac, Seibold, and Farris (2000) a	36	50	$F = .91$	.20	IN	M	3	A	Oth	Str	Both	2	Lab	Strx
Mulac et al. (2000) b	36	50	$F = .81$	.19	HE	M	3	A	Oth	Str	Both	2	Lab	Strx
Precht (2008)	183	183	$F = 0$	0	HE	F	n/a	A	Oth	Oth	Both	n/a	Oth	Unst
Reid, Keerie, and Palomares (2003) a	21	21	$F = 7.83$	.61	GEN	M	10	V	UG	Str	MG	2	Lab	Strx
Reid et al. (2003) b	21	21	$F = .11$	.07	GEN	M	10	V	UG	Str	MG	2	Lab	Strx
Rubin and Nelson (1983)	20	20	$p = 1$	0	GEN	M	n/a	A	Oth	Str	Both	2	Lab	Strx
Sayers and Sherblom (1987)	11	11	$t = .21$	.09	UN	W	n/a	V	Oth	Str	MG	2	Oth	Unst
Smeltzer and Watson (1986) a	36	36	$\chi^2 = 10.70$	.34	UN	M	75	V	Oth	Oth	2	4	Oth	Strx
Smeltzer and Watson (1986) b	36	36	$\chi^2 = 0.12$	−.06	TQ	M	75	V	Oth	Oth	2	4	Oth	Strx
Stiles et al. (1997) (Study 1)	19	19	$t = .65$	.15	GEN	M	6	V	UG	Str	MG	2	Lab	Unst
Stiles et al. (1997) (Study 2a)	7	6	$t = −.41$	−.23	GEN	M	5	V	UG	Str	SG	2	Lab	Unst
Stiles et al. (1997) (Study 2b)	11	11	$t = 1.43$	.43	GEN	M	5	V	UG	Str	MG	2	Lab	Unst
Stiles et al. (1997) (Study 3a)	12	12	$t = 2$	.58	GEN	M	30	A	Oth	Oth	MG	2	Lab	Unst
Stiles et al. (1997) (Study 3b)	12	12	$t = 1.14$	.33	GEN	M	30	A	Oth	Str	MG	2	Lab	Unst
Stiles et al. (1997) (Study 4)	33	33	$t = 2.63$	.46	GEN	M	15	A	UG	Oth	MG	2	Lab	Strx
Whitney (1991) a	10	10	M & SD	0	HE	W	15	A	UG	Str	MG	2	Lab	Strx
Whitney (1991) b	10	10	M & SD	0	IN	W	15	A	UG	Str	MG	2	Lab	Strx

Note. n/a = information not available; N (w) = number of women; N (m) = number of men; Stat value = statistical value; M & SD = effect size computed from reported means and standard deviations; d = aggregate effect size; Op def = operational definition; GEN = general tentative language; UN = expressions of uncertainty (disclaimers or qualifiers); HE = hedges; TQ = tag questions; IN = intensifiers; author gender = first author's gender; W = woman; M = man; Obs length = length of observation (in minutes); Record = method of recording interaction; A = audiotape only; V = video-tape; O = on-spot observations only; Sample = type of population sampled; UG = undergraduate students; Oth = other; Relation = relationship between participants; Str = strangers; Oth = other; Gender Comp = gender composition of dyad or group; SG = same gender; MG = mixed gender; both = same gender and mixed gender combined; Oth = other; Size = group size (number of participants observed interacting together); Setting = observational setting (lab or other); Activity = type of activity observed; Unst = unstructured; Strx = structured task or assigned conversational topic. Entries in *Author and Year* with different alphabetical subscripts refer to reports that included different samples, conditions, or measures of tentative speech. Entries with different alphabetical and numerical subscripts refer to different samples/conditions that additionally tested more than one measure of tentative speech.