

Perceptions of Gender and Femininity Based on Language: Implications for Transgender Communication Therapy

Language and Speech

1–19

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830914549084

las.sagepub.com**Adrienne B. Hancock**

The George Washington University, USA

Holly Wilder Stutts

The Lab School of Washington, USA

Annie Bass

Boston University, Boston

Abstract

Recent research presents a picture of diminishing gender differences in language. Two experiments examined whether language use can predict perceptions of gender and femininity; one included male and female speakers telling a personal narrative, the other also included male-to-female transgender speakers and analyzed an oral picture description. In each experiment, raters read transcribed samples before judging the gender and rating the femininity of the speaker. Only number of *T*-units, words per *T*-unit, dependent clauses per *T*-unit, and personal pronouns per *T*-unit emerged as different between gender groups. As none of the variables were strongly correlated with perceptual judgments, regression analysis was used to determine how combinations of linguistic variables predict female/feminine ratings. Results from these two studies demonstrate that gender-related differences in language use for these two contexts are limited, and that any relationship of language to perceptions of gender and femininity is complex and multivariate. This information calls into question the utility of training key language features in transgender communication therapy.

Keywords

Transgender, transsexual, gender perception, language

Corresponding author:

Adrienne B. Hancock, PhD, George Washington University, 2115 G Street, NW, Washington, DC 20052, USA.

Email: hancock@gwu.edu

Introduction

1.1 Gender differences in language

Prior to the 1960s, the field of sociolinguistics examined social roles with the assumption that research findings based on men applied to everyone (Kramer, 1977). However, with the rise of the feminist movement, this assumption was called into question, and sociolinguistics began to examine gender differences, in terms of real differences as well as differences that are widely believed to be true, regardless of their actual existence (Kramer, 1977). Robin Lakoff's (1973) foundational work "Language and Woman's Place" shaped the direction of research in gender language differences in the 1970s, and is still referenced in almost every gender language study to date. Based on her own insight, Lakoff identified the variables which she determined were representative of women's language, calling the combination of these variables the "female register." In terms of lexical differences, Lakoff describes women's use of more specific color terms (e.g., *mauve*, *lavender*), "weaker" particles (e.g., less offensive expletives such as *oh fudge*), and empty adjectives (e.g., *divine*, *cute*). Regarding syntax and suprasegmentals, Lakoff describes women's use of tag questions (a question added onto the end of a declarative indicating uncertainty, e.g., *They like ice cream, don't they?*), rising intonation in declarative answers to questions, and the use of more polite requests as opposed to commands. Soon after publishing this paper, Lakoff (1975) supplemented these characteristics of women's speech with additional variables of hedges (e.g., *sort of*, *kind of*, *maybe*), intensive use of the word *so* (e.g., *That dog is so cute*), and excessively correct grammar (Lakoff, 1975). Although she provides no empirical research regarding the perception or actual presence of these variables, Lakoff's (1973, 1975) assertions served to spark discussion regarding differences between the language of men and women, and set the stage for further research to be conducted.

In 1973 Hirschman presented what she called preliminary evidence of gender differences in language behavior, though they were not entirely consistent with Lakoff's concept of feminine register. She examined six mixed- and same-gender dyads of four college students during 10-minute paired discussions about controversial topics. She reported that the women used more fillers (e.g., *um*, *you know*) and personal references (e.g., *we*, *you*, *I*) than men used, fewer third person references (e.g., *he*, *they*) and interruptions, but equal number of qualifiers (what Lakoff called hedges, in addition to uncertainty verbs, e.g., *I think*, *I wonder*) (Hirschman, 1994). Crosby and Nyquist (1977) performed an empirical study of Lakoff's hypothesis, conducting three studies in which the "female register" linguistic markers were measured. The first study observed conversations between two speakers of the same sex, while the second and third observed same- and mixed-sex dyads at an information booth and police station, respectively. Results were inconsistent, finding that the female register was used significantly more in women's speech than men's speech when in same-sex conversation, but no significant difference between genders occurred in information booth interactions. The female register was used by men and women inquiring at the police station, confirming the author's hypothesis that feminine register was used by people based upon roles during conversation and not necessarily status or gender. Therefore, one's perceived role may influence his or her language use.

The 1980s and 1990s saw an expansion of research in gender language differences, and the Gender-Linked Language Effect (GLLE) was more thoroughly explored (Mulac, 2006). Statistical modeling analyses were used to relate the frequencies of linguistic variables to perceptions of gender, as well as to attributes of socio-intellectual status, aesthetic quality, and dynamism (Mulac & Lundell, 1986; Mulac, Wiemann, Widenmann, & Gibson, 1988). The models identified clusters of variables that, when weighted together accordingly, can discriminate between male and female

samples fairly accurately. For example, Mulac and Lundell's (1986) discriminant analyses of transcripts from a picture description task resulted in eight variables indicative of male speakers: impersonals, fillers, elliptical sentences, units (i.e., words or vocalized pauses), justifiers, references to pictures, geographical references, and spatial references; and nine more indicative of females: intensive adverbs, personal pronouns, negations, verbs of cognition, dependent clauses with subordinating conjunctions understood, oppositions, pauses, tag questions, and mean length of sentence. Another study (Mulac et al., 1988) investigated language used during a problem-solving task and resulted in a different model; it overlaps slightly with Mulac and Lundell's (1986) model but mostly consists of additional variables: interruptions, directives, and fillers/conjunctions starting the sentence were indicative of male participants and questions; justifiers, intensive adverbs, personal pronouns and adverbs starting the sentence were indicative of females. The only conflicting result in these two particular models is for justifiers; they were attributed to males by Mulac and Lundell (1986) but to females in the problem-solving task used by Mulac et al. (1988).

Variables measured and consequently included in models were not consistent in all studies, but generally authors concluded there was great overlap between male and female language, with some key features able to support the idea of a GLE in most contexts. Personality characteristics were hypothesized for some language clusters, allowing generalizations about females being "relatively unassertive, sensitive, formal and person-oriented... [while] males were relatively informal, concerned with holding the floor, and thing-oriented" (Mulac & Lundell, 1986, p. 96). Appendix A defines some of the commonly used variables and notes whether the variable was associated with males, females, or both in previous research.

Recent research presents a picture of diminishing gender differences in language use as earlier findings are not replicated and in some cases contradicted; this is possibly due to new contexts or changes over time in societal language norms. Furthermore, results from more context-based analyses indicate that several linguistic variables differ between men and women only in certain contexts, with differences changing between conversation and writing as well as according to the formality of the situation or characteristics of the communication partner (Hancock & Rubin, in press; Hannah & Murachver, 1999; Leaper & Ayres, 2007). In 2001, Thomson and Murachver found no difference between males and females in use of questions in e-mails, yet in a context of professional role-play Mulac, Seibold, and Farris (2000) found higher use of questions in men and more directives in women. Also contradictory to previous evidence, Mulac et al. (2000) found more references to emotion in men's language, and Mehl and Pennebaker (2003) found no significant difference in reference to emotion in conversation. Newman, Groom, Handelman, and Pennebaker (2008) found that number of words and mean sentence length did not vary between genders when considering all communication contexts or conversational context in isolation. For example, self-references were used more frequently by women in the more functionally applied conversational context.

Mulac and Lundell (1986) categorized males' picture descriptions as "relatively empirical (with a) thing-oriented perspective" whereas female discourse was characterized by "a general lack of assertion or, alternatively, a higher degree of interpersonal sensitivity" (p. 95). Therefore, it was somewhat surprising when naive observers reading transcripts were unable to identify the gender of the speaker but were able to discriminately rate socio-intellectual status, aesthetic quality, and dynamism. The language variables were used to develop strong statistical models of prediction for these three characteristics. Similarly, Mulac, Bradac, and Gibbons (2001) were able to relate gender to four dimensions of culture. Several variables were categorized as a feature of either males (direct, succinct, personal, instrumental) or females (indirect, elaborate, contextual, affective), but whether some of the same language variables could be used in a combination to develop a model to predict perceptions of speaker gender was not reported. Therefore, the utility of gender-related

language differences is unclear. As observed, differences are often minimal and mediated by other factors such as communication activity, relationship of speakers, group size and gender composition (Leaper & Ayres, 2007); therefore, the effect of language on listener/reader perception is questionable.

1.2 Clinical applications of gendered language research

Knowledge of how contemporary men and women use language and how the language used influences the way people perceive a speaker's gender can be applied by speech–language pathologists (SLPs). In recent decades, the scope of speech–language pathology services has grown to include communication therapy for transgender (TG) individuals. This therapy has evolved from simply raising vocal pitch and now often addresses several aspects of speech and language (Hancock & Garabedian, 2013). Goals are usually established based on normative gender data for that area. For some areas, particularly in language use, normative research cannot provide effective guidelines for treatment because of insufficient or inconsistent findings.

With the emergence of the TG population in SLPs' caseloads, the initial and most salient communicative change to address was pitch (Gelfer & Schofield, 2000; Wolfe, Ratusnik, Smith, & Northrup, 1990), and then also resonance and intonation (Carew, Dacakis, & Oates, 2007; Dacakis, 2002; Gelfer & Schofield, 2000; Hancock, Colton, & Douglas, 2014). Although less consistent, research has also examined the effects of volume, vocal quality, articulation, and rate on gender differentiation, and as a result these components may be included in communication therapy for TG speakers (Adler, Hirsh, & Mordaunt, 2012; Andrews & Schmidt, 1997; Fitzsimons, Sheahan, & Staunton, 2001; Oates & Dacakis, 1983; Van Borsel, Janssens, & De Bodt, 2009; Van Borsel & Maesschalck, 2008). Perhaps the least addressed communication variable in the TG communication literature is language. If language is targeted in therapy, SLPs rely on sociolinguistic literature for guidance. Hooper, Crutchley, and McCreedy (2012) recommend a variety of language therapy targets based on sociolinguistic research findings; however, many findings are of limited value because generalization is restricted to highly specific communicative contexts or particular age groups. In addition to research regarding observed gender differences, SLPs must rely on research regarding the influence of variables on perceived gender in order to guide their TG clients to be perceived as their desired gender. This balance of observed gender differences and influence on perception has proven useful in creating appropriate therapy targets for TG voice (e.g., pitch), but sufficient research has yet to be conducted to allow for this same balance when targeting language. Therefore, the influence of language on perception of gender is of great interest to TG individuals who work with SLPs to sound and communicate in such a way that they are perceived and accepted as their desired gender.

1.3 Purpose

The current study examined the nature of language differences between men and women, as well as how well particular language measures can predict gender and femininity perceptions. The first study included female and male groups telling a personal narrative; the second experiment added a group of male-to-female transgender women and analyzed a picture description sample. Inclusion of the transgender group provided an opportunity to measure language on a gender continuum, rather than in two discrete gender categories, and also directly examines the impact of language on perception of gender for transgender women. Two contexts were used because the literature suggests varying formality or constraints on language may impact conclusions about gender differences. Specifically, we aimed to:

1. Evaluate differences between gender groups for frequency of *T*-units (a measure of sentence length for spoken discourse) or any of the following variables (per *T*-unit): negations, dependent clauses, judgmental adjectives, progressive verbs, intensive adverbs, justifiers, hedges, qualifiers of time/quality, qualifiers of quantity, fillers, personal pronouns, self-references, words. (See Appendix A for operational definitions.)
2. Determine the value – if any – of the above variables for predicting gender judgments or femininity ratings by readers naive to the speaker's gender.

2 Method

2.1 Experiment 1

2.1.1 Materials. The purpose of presenting written transcription of spoken narratives was to control for confounding influence of speakers' voice characteristics (e.g., pitch) and disguise the speakers' gender. The Child Language Data Exchange System (CHILDES, Carnegie Mellon University, Pittsburgh, PA; MacWhinney, 2000) includes software to transcribe audio recordings of narratives (CHAT software) and code each sample for predetermined language variables (CLAN software). CHAT was selected because it was developed with the intention of providing a more accurate method for transcription and analysis compared to doing so by hand. CLAN is also beneficial in that it automates a portion of data analysis by performing functions such as frequency counts of specific words and coded variables.

A customized computer program was used to collect the perceptual data. This program displayed typed narratives, selection boxes for readers to make gender judgments for each speaker, and visual analog scales (which correlated to a point value of 0 to 1000) for the readers to rate each speaker's femininity. This program also displayed instructions and a practice narrative and rating. The program was linked with Microsoft Access software to record each reader's participant ID, age, and gender, as well as their ratings for each narrative.

2.1.2 Participants. A total of 40 narratives regarding a previous injury or illness were collected. Thirty narratives (19 by females, 21 by males) were downloaded from the TalkBank AphasiaBank database, an online databank of audio recordings and transcripts for language research (<http://talkbank.org/>). An additional 10 male narratives were collected by the researchers to supplement those downloaded from TalkBank. To mimic the protocol used by TalkBank researchers, the topic of a previous illness or injury was used in Experiment 1. This sample size was chosen to attempt to provide statistical power while remaining within the practical means of the researchers.

Selection criteria for all speakers (collected from TalkBank and otherwise) required that speakers self-report to be a minimum of 18 years old, have at least 12 years of completed education, be native English speakers, have no neurological, psychological, or cognitive disorders, have no hearing loss that affects speech, and have no expressive language disorder.

Forty readers (20 male, 20 female) participated in Experiment 1 by judging femininity and gender. The mean age of the readers was 22 years (ranging from 18–31), with a standard deviation of 2.75 years. Selection criteria required that all readers self-report to be a minimum of 18 years old, have a minimum of 12 years of education completed, be native English speakers, have no neurological, psychological, or cognitive disorders, have no current reading disability, and have no history of a language disorder. Readers were recruited through flyers posted on the George Washington University (GWU) campus, recruitment in courses taught at GWU, and word of mouth. They read narratives and provided a gender judgment and a femininity rating for each narrative. They received US\$15 for their participation. All materials were approved by The George Washington University's Institutional Review Board.

2.1.3 Procedures for collecting narratives. A total of 21 males (11 downloaded from TalkBank, 10 collected by the researchers) and 19 females (all downloaded from TalkBank) were asked to talk about an illness or injury. Of the narratives downloaded from TalkBank, 25 included audio recordings and transcriptions, while five included only audio recordings. For those that did not include transcriptions, researchers used the audio recordings to transcribe the narrative in CHAT (see section 2.1.4.: Procedures for transcribing narratives). Demographic speaker data was collected from TalkBank to ensure that all speakers met selection criteria.

Supplemental narratives were collected from individuals who were unfamiliar with the study. Each speaker completed a medical history form to ensure that selection criteria were met and signed an informed consent, and was asked to talk for three to five minutes about an illness or injury. This was recorded with a digital voice recorder and collected outside of the clinical setting.

2.1.4 Procedures for transcribing narratives. Audio recordings of narratives were transcribed using CHAT software. CHILDES provides a manual of transcription protocol, which was followed to ensure the accuracy of each transcription (MacWhinney, 2000). Using this protocol, running speech is segmented by pauses, and each segment is linked to audio that can be played back, for the entire language sample or for an individual segment. Additionally, this protocol allows for differential transcription of relevant speech components, such as fragments, abbreviated speech, lengthy pauses, and quotations.

In order to prevent biased ratings based on gender-indicative content rather than the language of the speaker, words that were clearly indicative of the gender of the speaker were changed. This included pronouns, names (e.g., *Sarah*), and titles (e.g., *husband/wife*) when referring to a spouse, and gender preferential sports (e.g., *football*), activities (e.g., *fishing*) or jobs (e.g., *construction*). Words were changed in a way that did not affect the presence of a language variable being measured in this study (e.g., *football* was changed to *ball* rather than *a game* so as not to change the number of words spoken).

Inter-rater reliability was assessed for narrative transcription. All narratives used in this study (those transcribed by TalkBank researchers as well as those transcribed by researchers for this study) were checked by a speech–language pathology graduate student, who checked the correspondence between the audio recordings of the narratives against the typed transcripts, and marked any discrepancies. Point-by-point agreement was 99.6%, and any discrepancies marked by the graduate student were examined by the first author to make the final determination.

2.1.5 Procedures for coding narratives. Narratives were coded using CLAN software and guidelines. Definitions for each linguistic variable were determined using previous research (e.g., Mulac et al., 2001), and were further clarified through discussion among researchers (see Appendix A).

Inter-rater and intra-rater reliability were assessed for the linguistic coding of transcribed narratives. Two researchers compared codes for the same 12 narratives (30% of all narratives) for all variables, with the exception of separation into *T*-units. *T*-units were coded by two researchers for every narrative and point-by-point agreement was originally 93%. Following discussion, 100% were agreed upon. Eight linguistic variables originally selected based on their occurrence in previous literature of gendered language were eliminated due to insufficient occurrences (i.e., questions, tag questions, directives, and oppositions) or inability to establish a reliable definition (i.e., ellipticals, sentence initial adverbs, references to emotion, and locatives). Point-by-point agreement for the remaining linguistic variables in Appendix A was 80% for the 12 transcripts (30% of all the transcripts) used to measure inter-rater reliability. Each researcher also re-coded eight (20%) of the narratives at least one week after initial coding. Point-by-point agreement was 90% and met the 80% criterion (Johnson & Pennybacker, 1993).

2.1.6 Procedures for collecting perceptual data. Each transcribed narrative was formatted to a readable paragraph, with punctuation determined based on the perceived intent of the speaker as well as readability. Dashes were placed after word fragments and repeated words. Commas were placed after pauses when the pause correlated with a grammatically appropriate comma, and wherever necessary to maintain readability of the narrative. Periods were placed at the end of a thought whether or not this correlated with a pause in speech, due to the tendency in spoken language to string together thoughts without pausing. This paragraph version of the narrative was entered into the computer program to be used by the readers.

Readers completed a questionnaire to ensure that selection criteria were met, and signed an informed consent form. Using the previously described custom computer program, they read brief instructions followed by a practice narrative, in which they read an exemplar paragraph created by the researchers and provided ratings, after which they were asked if they had any questions regarding their task. Participants then read a total of 48 narratives. Forty of these were from different speakers, and eight were repeated to assess intra-reader reliability. After reading a narrative, readers provided a gender judgment (whether they perceived the speaker to be male or female) and a femininity rating (a visual analog scale anchored by “feminine” and “masculine” anchor terms). Each reader received US\$15 for their participation in the study. Each participant took approximately one hour to complete the study.

To prevent a potential reading order effect, half of the readers received narratives 1–48 in consecutive order, while the other half received narratives 25–48 first, followed by 1–24. To assess intra-reader reliability, eight of the narratives (20%) were randomly selected to be repeated and distributed within the rest of the narratives. They were positioned semi-randomly, with the researchers only controlling to ensure that half of the repetitions were in narratives 1–24, and half were in narratives 25–48, and to ensure that the initial reading was not immediately followed by the repetition of that narrative.

2.2 Experiment 2

2.2.1 Participants. A total of 35 narratives were collected. Ten men, 12 women, and 13 transgender women (male-to-female; hereafter MtF) each described the same picture. Selection criteria for all speakers required that speakers self-report to be a minimum of 18 years old, have at least 12 years of completed education, be native English speakers, have no neurological, psychological, or cognitive disorders, have no hearing loss that affects speech, and have no expressive language disorder. The selection criteria were set due to the potential for each of these factors to differentiate the language of the speaker from typical adults. Speakers were given US\$10 for their participation in a larger study, of which this picture description task was a portion.

Forty readers (20 male, 20 female) participated in this study by judging femininity and gender. The mean age of the readers was 22 years (ranging from 18–31), with a standard deviation of 2.75 years. Selection criteria required that all readers be a minimum of 18 years old, have a minimum of 12 years of education completed, be native English speakers, self-report no neurological, psychological, or cognitive disorders, no current reading disability, and no history of a language disorder. Readers were recruited through recruitment in courses taught at George Washington University. They read narratives and provided a gender judgment and a femininity rating for each narrative. They were given US\$10 for their participation. All materials were approved by the institution’s IRB.

2.2.2 Procedures for collecting narratives. A total of 10 men, 12 women, and 13 transgender women (MtF) were asked to describe Norman Rockwell’s painting “The Waiting Room” for 20–30 seconds. Each speaker completed a medical history form to ensure that selection criteria were met and signed an informed consent form. Two of the narratives were omitted from rating because of their outlier

characteristics: one female speaker talked about a subject unrelated to the prompt, and one MtF speaker used unusual language (e.g., “ochre-colored bench”) and provided only a very short sample (i.e., 35 words). Thus, analyses are based upon 10 men, 11 women, and 12 MtF women.

2.2.3 Procedures for transcribing narratives. Audio recordings of all narratives were transcribed by the authors, using CLAN software. A speech–language pathology graduate student checked the correspondence between the audio recordings of the narratives and the typed transcripts, and marked any discrepancies. Discrepancies marked by the graduate student were rechecked by the first author to make the final determination.

2.2.4 Procedures for coding narratives. Narratives were coded using CLAN software. The same linguistic variables from Experiment 1 were used. Inter-rater and intra-rater reliability were assessed for the coding of transcribed narratives. A sample of six narratives (18%) were coded for *T*-units by two researchers and point-by-point agreement was originally 91.4%, with 100% agreed upon following discussion. One researcher then coded *T*-units for the remaining 33 narratives, and a second checked them; 98% reliability was obtained, with 100% agreed upon after discussion. Six (18%) of the transcripts were coded for the other linguistic variables by two people for measurement of inter-rater reliability. Point-by-point agreement for linguistic variables was 87%, with 100% agreed upon after discussion. In order to achieve consistency, the definitions of the linguistic variables were clarified further. Appendix A is the final definitions used to code the data included in this paper. Point-by-point agreement was then 92%.

2.2.5 Procedures for collecting perceptual data. Each transcribed narrative was formatted to a readable paragraph form, with punctuation determined based on the intent of the speaker as well as readability. This paragraph version of the narrative was entered into the computer program to be used by the readers. Readers completed a questionnaire to ensure that selection criteria were met, and signed an informed consent form. Using the previously described computer program, they read brief instructions followed by a practice narrative, in which they read an exemplar paragraph and provided ratings, after which they were asked if they had any questions regarding their task. Participants then read a total of 40 narratives. Thirty-three of these were from different speakers, and seven were repeated to assess intra-reader reliability.

After reading a narrative, readers provided a gender judgment (whether they perceived the speaker to be male or female) and a femininity rating (a sliding scale anchored by “feminine” and “masculine” terms). Each reader received US\$10 for their participation in the study. Each participant took approximately 30 minutes to complete the study.

To prevent a potential reading order effect, half of the readers received narratives 1–40 in consecutive order, while the other half received narratives 21–40 first, followed by 1–20. To assess intra-reader reliability, six of the narratives (18%) were randomly selected to be repeated and distributed within the rest of the narratives. They were positioned semi-randomly, with the researchers only controlling to ensure that half of the repetitions were in narratives 1–20, and half were in narratives 21–40, and to ensure that the initial reading was not immediately followed by the repetition of that narrative. The practice paragraph was repeated at the end.

3 Results

3.1 Group comparisons

A post-hoc power analysis using $n = 40$, $\alpha = .05$, and groups = 2 (e.g., Experiment 1) revealed .69 power for large effect sizes, .34 power for medium effect sizes, and .09 for small effect sizes.

Table 1. Means (and standard deviations) of language variables for Experiment 1 by gender group.

| Linguistic variable | Males (n=21) | Females (n=19) | F | p | eta ² |
|----------------------------|--------------|----------------|-------|------|------------------|
| T-units | 29 (15) | 16 (11) | 8.67* | .006 | .19 ^a |
| Negations | .18 (.11) | .15 (.14) | .49 | .488 | |
| Dependent clauses | .29 (.17) | .16 (.08) | 7.36* | .010 | .16 ^a |
| Judgmental adjectives | .14 (.12) | .16 (.08) | .47 | .496 | |
| Progressive verbs | .24 (.15) | .21 (.12) | .57 | .455 | |
| Intensive adverb | .20 (.16) | .22 (.17) | .21 | .649 | |
| Justifier | .05 (.04) | .03 (.04) | 3.85 | .057 | |
| Hedges | .18 (.10) | .14 (.13) | 1.24 | .272 | |
| Qualifiers of time/quality | .12 (.11) | .09 (.09) | .75 | .392 | |
| Qualifiers of quantity | .34 (.16) | .31 (.13) | .46 | .502 | |
| Fillers | .49 (.24) | .50 (.44) | .01 | .927 | |
| Personal pronouns | 1.42 (.36) | 1.21 (.24) | 4.49* | .041 | .10 ^b |
| Self-references | 1.03 (.32) | 1.15 (.51) | .77 | .386 | |
| Words per T-unit | 11.6 (2.1) | 9.70 (2.8) | 6.21* | .017 | .14 ^a |

Note: variables are analyzed in a ratio of variable per T-unit.

*= Statistically significant at the $p < .05$ level.

^a = large effect size ^b = medium effect size.

Another power analysis using $n = 33$, $\alpha = .05$, and groups = 3 (e.g., Experiment 2) revealed less than .50 power for all effect sizes.

A one-way ANOVA revealed statistically significant differences between males and females in Experiment 1 for number of T-units, $F(1,39) = 8.67$, $p = .006$; words per T-unit, $F(1,39) = 6.21$, $p = .017$; dependent clauses, $F(1,39) = 7.36$, $p = .010$; and personal pronouns, $F(1,39) = 4.49$, $p = .041$; with females always having fewer than males. The effect sizes of the former three variables were large (i.e., $\eta^2 > .14$). See Table 1.

For Experiment 2, an ANOVA revealed significant between-group difference for T-units, $F(2,32) = 3.36$, $p = .048$; words per T-unit, $F(2,32) = 4.57$, $p = .019$; and dependent clauses, $F(2,32) = 3.86$, $p = .032$. Bonferroni post-hoc comparison tests revealed a significant difference between the female and the MtF group for the significant variables: T-units ($p = .045$), words per T-unit ($p = .017$), dependent clauses ($p = .050$). For each of these variables, males were not significantly different from either group; the mean of the male group was between the MtF and female group means. MtF speakers used the least T-units but the most words per T-unit and dependent clauses. These three variables had large effect sizes (i.e., $\eta^2 > .14$). See Table 2.

3.2 Prediction of gender perception

3.2.1 Experiment 1. Fifty-three percent (21/40) of the gender judgments were accurate. Seventy-one percent (15/21) of males' samples were judged to be male (three were undecided) and 31% (6/19) of the females' narratives were judged to be female. All 14 linguistic variables were first entered into a multiple linear regression model to describe how the variables predicted the gender judgment score given by raters. The fit of the model was not statistically significant, $F(14, 25) = 1.24$, $p = .31$. Therefore, forward, backward, and step-wise regression modeling was used to select variables to predict gender judgment. The model with the best fit indicated the sample was more likely to be judged as a female's with an *increase* in dependent clauses, justifiers, qualifiers of quantity, fillers, and personal pronouns and a *decrease* in progressive verbs, qualifiers of time and quality, and words per T-unit, $F(8,31) = 2.29$, $R^2 = .37$, $p = .047$.

Table 2. Means (and standard deviations) of language variables for Experiment 2 by gender group.

| Linguistic variable | Males (n=10) | Females (n=11) | MtF (n=12) | F | p | eta ² |
|----------------------------|--------------|----------------|------------|-------|------|------------------|
| T-units | 5.4 (1.8) | 6.5 (1.6) | 4.7 (1.3) | 3.36* | .045 | .18 ^a |
| Negations | 0.4 (.97) | 0 | 0.1 (.3) | 1.49 | .241 | |
| Dependent clauses | 2.5 (1.3) | 1.4 (1.1) | 2.6 (1.1) | 3.86* | .032 | .20 ^a |
| Judgmental adjectives | 0.5 (.5) | 0.1 (.3) | 0.3 (.7) | 1.66 | .207 | |
| Progressive verbs | 2.6 (1.6) | 1.3 (1.1) | 2.1 (1.4) | 2.60 | .091 | |
| Intensive adverb | 0.4 (.5) | 1.3 (1.1) | 0.8 (.8) | 2.68 | .090 | |
| Justifier | 1.0 (.8) | 1.0 (.8) | 1.5 (1.7) | .64 | .536 | |
| Hedges | 4.4 (3.2) | 4.4 (2.4) | 3.1 (2.2) | .95 | .399 | |
| Qualifiers of time/quality | .5 (1.1) | 0.7 (.7) | 0.3 (.5) | .77 | .471 | |
| Qualifiers of quantity | 3.3 (2.1) | 3.9 (1.6) | 2.6 (1.7) | 1.59 | .222 | |
| Fillers | 1.8 (1.4) | 1.9 (1.6) | 3.3 (2.8) | 1.91 | .166 | |
| Personal pronouns | 3.4 (2.3) | 3.3 (1.7) | 2.3 (1.3) | 1.19 | .318 | |
| Self-references | 0.1 (.3) | 0.3 (.5) | 0.5 (.5) | 2.19 | .130 | |
| Words per T-unit | 12.7 (4.3) | 9.8 (2.6) | 14.2 (3.5) | 4.57* | .019 | .23 ^a |

Note: variables are analyzed in a ratio of variable per T-unit.

* = Statistically significant at the $p < .05$ level.

^a = large effect size.

All 14 linguistic variables were then entered into a multiple linear regression model to describe how the variables predicted the femininity score given by raters. The fit of the model was not significant, $F(14,25) = 1.19, p = .338$. Therefore, forward, backward, and step-wise regression modeling was used to select variables to predict femininity rating. The model with the best fit included the same eight variables that were included in the model to predict gender judgment (i.e., increase in femininity rating with an *increase* in dependent clauses, justifiers, qualifiers of quantity, fillers, and personal pronouns and a *decrease* in progressive verbs, qualifiers of time and quality, and words per T-unit), $F(8,31) = 2.17, R^2 = .36, p = .059$.

3.2.2 Experiment 2. Forty-five percent (15/33) of the gender judgments were accurate. Fifty percent (5/10) of males' samples were judged to be male, 36% (4/11) of the females' narratives were judged to be female, and 50% (6/12) of the MtF speaker samples were judged to be female. The speaker's gender and measures of all linguistic variables were entered into a multiple linear regression model to describe how the variables predicted the gender judgment score given by raters. The fit of this model was not statistically significant, $F(14,18) = 2.18, p = .06$. Total T-units had a strong linear correlation with words per T-unit ($r = -.88$) and was therefore removed to avoid redundancy, but the resulting model was still not significant, $F(13,19) = 2.35, p = .05$. Therefore, forward, backward, and step-wise regression modeling was used to select variables to predict gender judgment. The model with the best fit indicated the sample was more likely to be judged as a female's with an *increase* in dependent clauses, intensive adverbs, qualifiers of quantity, fillers, personal pronouns, and a *decrease* in progressive verbs, hedges, and self-references, $F(8,24) = 4.20, R^2 = .58, p = .003$.

The speaker's gender and measures of all linguistic variables were entered into a multiple linear regression model to describe how the variables predicted the femininity score given by raters. The fit of this model was not statistically significant, $F(14,18) = 1.77, p = .126$. Total T-units had a

strong linear correlation with words per *T*-unit ($r = -.88$) and was therefore removed, but the resulting model was still not significant, $F(13,19) = 1.93, p = .093$. Therefore, forward, backward, and step-wise regression modeling was used to select variables to predict femininity score. The model with the best fit indicated femininity score increased with an *increase* in dependent clauses, intensive adverbs, qualifiers of quantity, and personal pronouns and a *decrease* in progressive verbs, hedges, and self-references, $F(7,25) = 3.74, R^2 = .51, p = .007$. With the exception of fillers in the gender judgment model, the models for gender judgment and femininity rating in Experiment 2 included the same variables.

3.3 Summary

Four of the 14 variables differentiated males from females (*T*-units, words per *T*-unit, dependent clauses, personal pronouns). Unexpectedly, the MtF speakers tended to be more distinct from the females than even the males were; this was to a significant extent for number of *T*-units, words per *T*-unit, and dependent clauses.

Raters' judgment of a speaker's gender based on reading a transcript of a spoken narrative was at chance levels in both experiments (53% and 45%). Regression models for perception of gender and femininity were similar within each experiment. Four variables were common to all four models: dependent clauses, progressive verbs, qualifiers of quantity, and personal pronouns. In narrative context (Experiment 1), justifiers, qualifiers of time/quality, and words per *T*-unit were also in the models predicting gender and femininity; whereas in a picture description context (Experiment 2) intensive adverbs, hedges, and self-references were the additional variables. No models were strengthened by total number of *T*-units, negations, or judgmental adjectives.

4 Discussion

4.1 Gendered language

Overall, our studies do not provide strong evidence for language differences between males and females. Compared to males, females used significantly lower numbers of four of the 14 variables measured, and these differences occurred in only one of the two contexts (personal narratives). Our female subjects used significantly fewer *T*-units, words, dependent clauses, and personal pronouns than males during picture description.

Although females were expected to talk more (i.e., use more *T*-units and more words per *T*-unit) and more elaborately (i.e., use more dependent clauses) than males, they consistently – and significantly in Experiment 1 – used fewer of each of these variables. Considering females are characterized as preferring to talk about personal matters and affiliate with the listener, it is surprising that a more “feminine style” did not occur in a personal narrative (Mulac & Lundell, 1986; Mulac et al., 2001). Perhaps the topic of the personal narrative (i.e., an injury), was masculine and therefore influenced use of gendered language (Palomares, 2009).

Although females unexpectedly used *fewer* dependent clauses and personal pronouns than men did in Experiment 1, these variables were included in the perceptual models with positive coefficients – indicating an *increase* in use would actually increase perception of femininity. This suggests a mismatch between actual and stereotypical gender differences. For practical applications (e.g., transgender communication strategies), the impact of language use on perception of gender is of utmost value.

Table 3. Unstandardized coefficients for variables included in final regression models for gender rating and femininity score outcomes in each experiment.

| | Experiment 1 | | Experiment 2 | |
|----------------------------|---------------|------------------|---------------|------------------|
| | Gender rating | Femininity score | Gender rating | Femininity score |
| Negations | | | | |
| Dependent clauses | .73 | 347.52 | .05 | 24 |
| Judgmental adjectives | | | | |
| Progressive verbs | -.33 | -94.60 | -.04 | -23.26 |
| Intensive adverb | | | .06 | 27.69 |
| Justifier | 1.01 | 508.61 | | |
| Hedges | | | -.03 | -15.94 |
| Qualifiers of time/quality | -.51 | -173.28 | | |
| Qualifiers of quantity | .37 | 116.50 | .05 | 24.77 |
| Fillers | .16 | 98.32 | -.02 | |
| Personal pronouns | .12 | 70.95 | .04 | 16.43 |
| Self-references | | | -.06 | -74.76 |
| Words per T-unit | -.03 | -17.04 | | |

Notes: Variables are analyzed in a ratio of variable per T-unit. Mean gender ratings for each sample could range 0–1, with 1 as more female. Mean femininity score for each sample could range 0–1000, with 1000 as more feminine.

4.2 Language and perception of gender

Gender prediction accuracy was near chance in both experiments. None of the individual variables alone were robust predictors of perceived gender or femininity. Only through regression modeling, a statistical method for examining the value of using combinations of variables to predict an outcome, were good prediction models found. As expected, models for predicting gender and femininity were similar *within* each experiment in that they included the same eight linguistic variables with the same coefficient signs. However, the eight variables were not consistent *across* experiments (i.e., different language tasks). The variation in linguistic variables included in the models for these experiments affirms the importance of considering context and associated influences when examining how language could predict perceptions (Crosby & Nyquist, 1977; Hancock & Rubin, in press; Hannah & Murachver, 1999).

An important note about interpreting regression models is warranted. In order to influence these outcomes (in this study, perceptions), one must address each linguistic variable in the model by using it more or less frequently, as determined by the unstandardized coefficient (see Table 3). The magnitude of the coefficient indicates to what extent changing the frequency of the variable will change perception, whereas the sign of the coefficient indicates whether an increase or decrease of the variable's frequency will increase the likelihood of the outcome. It would be inappropriate, for example, to say that if all four models include dependent clauses then a person who is male-to-female transgender should use more dependent clauses in order to pass as female or more feminine. Instead, all variables included in the model for the context of interest would need to be changed to the direction and extent indicated by the coefficient. With this understanding in mind, we proceed cautiously to examine the linguistic variables in light of previous literature. The direction of the coefficients (i.e., whether an increase in the variable is associated with an increase or decrease in femininity score) was not always as predicted based on previous literature.

4.3 Variables associated with increased perception of femininity

First, we consider the variables with a positive coefficient, indicating an increase in female/feminine scores with *increasing* use of these variables: dependent clauses, personal pronouns, intensive adverbs, qualifiers of quantity, and justifiers. The first three are consistent with the literature and the conceptualization of female speech to be complex, descriptive, and affiliative (Hirschman, 1994; Mulac & Lundell, 1986).

All the perceptual models in this study included positive coefficients for qualifiers of quantity. This is in contrast to studies that examined written communication and attributed qualifiers of quantity to men because they were used in a direct speaking style focused on external attributes of objects, as opposed to a more affective speaking style focusing on internal feelings (Mulac, Studley, & Blau, 1990; Mulac et al., 2001).

Greater use of justifiers associated with an increase in perception of femininity was somewhat surprising, given Mulac and Lundell's (1986) report of justifier use as indicative of male speakers in their multiple regression analysis of picture descriptions. However, justifiers have been used by females more than males in other contexts (e.g., problem-solving tasks, Mulac et al., 1988) and people may perceive speakers who justify to be more submissive and therefore associate them with a female stereotype.

4.4 Variables associated with decreased perception of femininity

Next, the variables with a negative coefficient, indicating an increase in female/feminine scores when frequency *decreased* were: progressive verbs, self-references, hedges, qualifiers of time/quality, and words per *T*-unit. Previous literature reports females using fewer progressive verbs in a verbal context, compared to men (Mulac & Lundell, 1994; Mulac, Lundell, & Bradac, 1986); however, the other variables with negative coefficients were not necessarily expected to contribute to a masculine perception.

In Experiment 2, using self-references decreased the perceived femininity. This is consistent with Mulac et al.'s (2001) assertion that "I" references are associated with male culture and "reflect an ego-centric orientation" (p. 144). However, Newman et al. (2008) found women to use more self-references than men in conversational contexts and explained this as a reflection of females' tendency to attend to personal aspects of conversation. However, those findings may be dependent upon a conversation partner (Hancock & Rubin, in press), which would not be an issue in the current study's contexts of personal narrative and picture description.

There is some discrepancy in the literature about hedges. According to Lakoff and theories suggesting women's language reflects a submissive status, hedges or uncertainty words are more expected from women (Lakoff, 1975). Crosby and Nyquist (1977) found women more likely than men to use hedges when requesting information. However, several studies, including this one, found no distinction between genders for number of hedges, sometimes called uncertainty modifiers (e.g., Hirschman, 1994; Mulac & Lundell, 1986; Mulac & Lundell, 1994; Mulac et al., 1986; Mulac et al., 1988). To our knowledge, no previous studies have indicated that hedges would be associated with gender judgment of male and low femininity rating, as it was in Experiment 2.

Qualifiers were examined in previous gender language research in terms of the subcategories of qualifiers of certainty or references to quantity (Mulac & Lundell, 1994). However, qualifiers can also include those qualifying time and relative quality, which are unexamined to date. As a result, the current study included qualifiers of time and relative quality to obtain information about gender differentiation as reflected in all qualifiers. This has proven interesting because qualifiers of quantity were associated with female/feminine scores while qualifiers of time and quality were associated

with male/masculine scores (in Experiment 2). However, in future studies it may be valuable to further separate qualifiers of time and quality into “relative” and “absolute” categories, because relative qualifiers, such as *some*, *usually*, or *nearly*, indicate uncertainty, while absolute qualifiers, such as *all*, *always*, or *never* indicate certainty. Such differentiation could provide insight on the occurrence and influence of qualifiers, and contribute to the discussion of whether relative qualifiers might be considered hedges, and whether they are more frequently used by men or women.

4.5 Variables with neutral effect on perception of femininity

Finally, because no models were strengthened by the total number of judgmental adjectives or negations, we can conclude that these variables were not valuable in these experiments, even in combination with other language variables, for predicting these perceptual outcomes. This is not to say they may never be of interest. Previous studies concluding that judgmental adjectives or phrases are predominately used by males measured written communication (Mulac & Lundell, 1994; Mulac et al., 1990), which may explain why judgmental adjectives were not significant in this study of oral language. Also, the finding that negations did not contribute to any models of perception may be explained by the fact that few negations were used at all in these samples; none of the 11 females in Experiment 2 used even one negation, despite previous findings that females use more negations than males use (Mulac & Lundell, 1986).

Fillers were associated with an increase in female/femininity scores in models using Experiment 1 results, but a small negative coefficient was required in the gender judgment model of Experiment 2. This may not be surprising considering the literature is also mixed regarding how this variable is used by men and women (Hirshman, 1994; Mulac & Lundell, 1986; Mulac & Lundell, 1994; Mulac et al., 1986; Mulac et al., 1990). Studies have attributed fillers to females, as Experiment 1 does, but they involved the writing of 4th graders (Mulac et al., 1990), the public speaking of university students (Mulac et al., 1986), and a small study comparing two females and two males in conversational contexts (Hirschman, 1994). Experiment 2 and Mulac and Lundell (1986) both indicate males use more fillers – and they both used description tasks. Thus, the tendency for a particular gender to use more fillers is another clear example of the effect of context.

5 Limitations

There was a low frequency of occurrence of many linguistic variables. This may have been due to the narrowness of the topic as well as the length of the language sample. Although longer or additional samples may increase the statistical power of the multiple linear model, the feasibility of meeting these requirements was beyond the scope of the design of the current study in that it would be unreasonable for the readers to evaluate 40 samples longer than those used in this study, or more than 40 samples. Future studies may be designed to accommodate this potential limitation by including several rater groups.

Internal validity may also be affected by low statistical power, low percentages of females judged as females, and constraints of the linguistic variables selected for measurement. Variables which were not reported previously occurred minimally in the current study’s language samples (i.e., questions, tag questions, directives, and oppositions), or were too complex to code reliably (i.e., ellipticals, sentence initial adverbs, references to emotion, and locatives), were not included in the analysis. It is possible that linguistic variables not measured account for additional gender language differences or would increase the percentage of samples correctly judged to be from a female speaker. External validity is constrained by the culture of the speakers and raters in this sample. For example, raters were largely university students and staff.

Over the last several decades, gender-based language patterns have not been robust, to say the least. One factor contributing to the lack of compelling evidence is that linguistic variables examined vary among studies and the actual coded features and interpretations often differ. Methodologies have changed with the continual development of technology to use for data collection and analysis. Furthermore, language use itself may have evolved, further complicating the current application of available research.

6 Implications for transgender speakers and future directions

The primary purpose of including an MtF group in the second experiment was to gain insight into the clinical utility of any differences between male and female language use. It was hypothesized that, if there were gender differences, the MtF group would adopt language more similar to the female group, with the average scores of the MtF falling between the male and female group scores. Instead, they did not statistically differ from males and in fact often the MtF group mean was closer to the male mean than the female mean. Perhaps not surprisingly, with few observable differences between the language of male and female gender groups, accuracy of gender prediction after reading transcripts was near chance. The lack of measured and perceptible differences calls into question the utility of training key language features in transgender communication therapy.

However, other literature has found greater gender differences and also found perception of gender to be influenced by language. Palomares (2004, 2009) recently suggested that perception of gender may be influenced by the recipient's stereotypes and gender identity schema in addition to language choices made by the speaker. Therefore, continued research in this area could investigate what the speaker could do in order to capitalize upon (or minimize) the recipient's gender schema so that language choices could be used to influence perception of gender. This would be particularly valuable for the written communication of transgender people, such as email (Palomares, 2004).

7 Conclusions

These two experiments demonstrate that gender-related differences in language use are limited, and that any relationship of language to perceptions of gender and femininity is complex and multivariate. Specific linguistic variables may be relatively more predictive than others, but none emerge as clearly useful for influencing how people perceive gender or femininity.

Acknowledgements

The authors wish to acknowledge Amy Lopez for her contributions to linguistic analysis and Xian Sun and Liyi Jia for assistance with statistical analysis.

Funding

The research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Adler, R. K., Hirsch, S., & Mordaunt, M. (2012). *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide* (2nd ed.). San Diego, CA: Plural Pub.
- Andrews, M. L., & Schmidt, C. P. (1997). Gender presentation: Perceptual and acoustical analyses of voice. *Journal of Voice*, 11(3), 307–313.

- Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice*, *21*(5), 591–603.
- Crosby, F., & Nyquist, L. (1977). The female register: An empirical study of Lakoff's hypothesis. *Language in Society*, *6*, 313–322.
- Dacakis, G. (2002). The role of voice therapy in transsexuals. *Current Opinion in Otolaryngology and Head and Neck Surgery*, *10*, 173–177.
- Fitzsimons, M., Sheahan, N., & Staunton, H. (2001). Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies. *Brain and Language*, *78*(1), 94–108.
- Gelfer, M. P., & Schofield, K. J. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice*, *14*(1), 22–33.
- Hancock, A. B., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice*, *28*(2), 203–209.
- Hancock, A. B., & Garabedian, L. (2013). Transgender voice and communication treatment: A retrospective chart review of 25 cases. *International Journal of Language and Communication Disorders*, *48*(1), 54–65.
- Hancock, A. B., & Rubin, B. A. (in press). Influence of communication partner's gender on language. *Journal of Language and Social Psychology*. Advance online publication. doi: 10.1177/0261927X14533197.
- Hannah, A., & Murachver, T. (1999). Gender and conversational style as predictors of conversational behavior. *Journal of Language and Social Psychology*, *18*, 153–174.
- Hirschman, L. (1994). Female–male differences in conversational interaction. *Language in Society*, *23*, 427–442.
- Hooper, C., Crutchley, S., & McCready, V. (2012). Syntax and semantics: A menu of communicative choices. In R. Adler, S. Hirsch, & M. Mordaunt (Eds.), *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide* (2nd ed., pp. 397–217). San Diego, CA: Plural Pub.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. (Research Rep. No. 3). Champaign, IL: National council of Teachers of English.
- Johnson, J. M., & Pennybacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kramer, C. (1977). Perceptions of female and male speech. *Language and Speech*, *20*, 151–161.
- Lakoff, R. T. (1973). Language and woman's place. *Language in Society*, *2*, 45–80.
- Lakoff, R. T. (1975). *Language and woman's place*. New York, NY: Harper & Row.
- Leeper, C., & Ayres, 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*(4), 328–363.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mehl, M. R., & Pennebaker, J.W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*(4), 857–870.
- Mulac, A. (2006). The gender-linked language effect: Do language differences really make a difference? In K. Dindia & D. Canary (Eds.), *Sex differences and similarities in communication* (2nd ed., pp. 219–239). Mahwah, NJ: Erlbaum.
- Mulac, A., Bradac, J., & Gibbons, P. (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, A., & Lundell, T. L. (1986). Linguistic contributors to the gender-linked language effect. *Journal of Language and Social Psychology*, *5*(2), 81–101.
- Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse – Multivariate tests of language effects. *Language & Communication*, *14*(3), 299–309.
- 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*(2), 115–129.

- Mulac, A., Seibold, D., & Farris, J. (2000). Female and male managers' and professionals' criticism giving – Differences in language use and effects. *Journal of Language and Social Psychology, 19*(4), 389–415.
- Mulac, A., Studley, L.B., & Blau, S. (1990). The gender-linked language effect in primary and secondary students' impromptu essays. *Sex Roles, 23*(9/10), 439–469.
- 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: A Multidisciplinary Journal, 45*(3), 211–236.
- Oates, J. M., & Dacakis, G. (1983). Speech pathology considerations in the management of transsexualism: A review. *British Journal of Disorders of Communication, 18*, 139–151.
- Palomares, N. A. (2004). Gender schematicity, gender identity salience, and gender-linked language use. *Human Communication Research, 30*, 566–588.
- 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 counter-stereotypically as a function of gender salience. *Communication Research, 36*, 538–560.
- Shadden, B. (1998). Obtaining the discourse sample. In Cherney, Shadden, & Coelho (Eds.), *Analyzing discourse in communicatively impaired adults* (pp. 9–34). Gaithersburg, MD: Aspen Publishers, Inc.
- Thomson, R., & Murachver, T. (2001). Predicting gender from electronic discourse. *The British Journal of Social Psychology, 40*, 193–208.
- Van Borsel, J., Janssens, J., & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: A perceptual approach. *Journal of Voice, 23*(3), 291–294.
- Van Borsel, J., & Maesschalck, D. (2008). Speech rate in males, females, and male-to-female transsexuals. *Clinical Linguistics & Phonetics, 22*(9), 679–685.
- Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech & Hearing Disorders, 55*, 43–50.

Appendix A

Definitions and citations for linguistic variables

- 1) Narrative level
 - a) *T-Unit*: one main clause plus any subordinate clauses or nonclausal structures attached to or embedded in the main clause. An index of syntactic complexity and useful as a common measurement base for other analyses (Hunt, 1965; Shadden 1998).
 - b) *Words per T-unit*: the number of total words divided by the number of total *T*-units in a language sample. (Mulac & Lundell, 1986, F+; Mulac et al., 1986, F+). This is sometimes termed mean length of utterance.
 - c) *Negation* (NO): a statement of what something is not. (Mulac & Lundell, 1986, F+; Mulac et al., 1986, F+.) This can consist of:
 - i. Turning an affirmative statement into its opposite denial (e.g., You don't feel like looking, I am not the winner).
 - ii. Negating a noun or verb using a negative adjective (e.g., There is no winner), a negative pronoun (e.g., Nobody is the winner), or a negative adverb (e.g., I never was the winner).
- 2) Phrase level
 - a) *Dependent clause*: a phrase that contains a subject and verb but cannot stand alone as a full sentence. A dependent clause usually begins with a subordinating conjunction

(e.g., that, which, who, whose, when, because, even though, whether, while) (Mulac & Lundell, 1986, F+). Dependent clauses can appear in various forms:

- i. Relative pronouns (e.g., I liked the dog who lived next door, I saw the man whose smile annoyed me, She had a pimple which wouldn't go away).
 - ii. Adverbial clauses (e.g., Be careful when leaving the house).
 - iii. Non-finite clauses (e.g., I wanted to be loved).
- b) *Progressive verb*: verb presented in the *-ing* form (Mulac et al., 1986, M+) (e.g., She was running).
- 3) Word level: semantic (meaning)
- a) *Justifier*: a reason is given for a previous statement (Mulac & Lundell, 1986, M+; Mulac et al., 1988, F+) (e.g., I left because I was hungry).
 - b) *Hedges*: (Crosby & Nyquist, 1977, F+)
 - i. Modifiers that indicate a lack of confidence in, or diminished assuredness of, the statement. A word or phrase that changes how absolute or certain a statement is (e.g., sort of, somewhat, kind of, probably, about).
 - ii. Verb phrase indicating lack of certainty or assuredness (e.g., I wonder if, it looks to be, I think, I guess, it seems to be).
 - c) *Qualifier of time or relative quality*
 - i. Qualifier of time indicates when or how often something occurs (e.g., occasionally, usually, sometimes, always, never, recently).
 - ii. Qualifier of relative quality compares the item at hand to something else, on a scale of degree (e.g., happier, coolest, worst nightmare, more satisfied, least hungry).
 - d) *Qualifiers of quantity*: reference to amount or quantity. Answers the questions *How much? How many?* Includes nominal or ordinal numbered quantities and qualifiers of quantity (Mulac & Lundell, 1994, M+; Mulac et al., 1990, M+) (e.g., all, most, none, each, any, some, first, below 32, 6–8 weeks).
- 4) Word level: syntactic (function)
- a) *Intensive adverb*: expresses how complete a quality is. As with all adverbs, must modify a verb, adjective, phrase, clause, or another adverb (Mulac & Lundell, 1986, F+; Mulac et al., 1986, F+; Mulac et al., 1986, M+; Mulac et al., 1988, F+) (e.g., very, really, quite, entirely, a little, a bit, pretty, more).
 - b) *Judgmental adjective*: rather than purely describing, judgmental adjectives evaluate *and comment on* the item under discussion. Many judgmental adjectives can be easily categorized as either good or bad (Mulac & Lundell, 1994, M+; Mulac et al., 1986, M+; Mulac et al., 1990, M+) (e.g., funny, horrible, annoying, nice, bothersome, ugly, fancy, cosmopolitan, old; NOT tired, bright, blurry, loud).
 - i. “He’s scared” is not a judgmental adjective, but “he’s scary” is.
 - c) *Personal pronoun*: word that refers to a person, place, thing, or idea (Mulac & Lundell, 1986, F+; Mulac et al., 1988, F+). Includes subject pronouns (e.g., I, you, he, she, it, we, they) and object pronouns (e.g., me, you, him, her, it, us, them).

- d) *Self-reference*: First person singular pronoun (e.g., I, me, my, mine, myself) (Mulac et al., 2001, M+; Newman et al., 2008, F+).
- e) *Filler*: Word or phrase used without apparent semantic intent. Repetitions of fillers are counted independently (Mulac et al., 1986, F+; Mulac & Lundell, 1986, M+; Mulac et al., 1988, M+) (e.g., like, er, I you know uh um went to bed).

Note: F+ study results attributed this variable to females; M+ study results attributed this variable to males.