

Automated Analysis of Spoken Language: A Gender Comparison

Terry Hallett and James Steiger

University of Akron

Hallett@uakron.edu steiger@uakron.edu

Abstract

We studied the number of words spoken by adult males versus females throughout a six-hour day and during three structured monologues. The six-hour samples were captured and analyzed using an automated speech monitoring and assessment system. The three monologues required different language tasks, and analyses of syntactic and semantic complexity were performed for each. There were no significant gender differences except during a reminiscent monologue when males spoke significantly more words and sentences than females. These results conflict with past research and popular (mis)conceptions.

Introduction

Gender difference studies have been reported in the psychological and linguistic literature regarding behavior, attitudes, and other sociological and psychological characteristics [1]. Researchers have documented differences in “genderlect,” a classic but controversial term referring to a variety of linguistic features distinguishing gender [2]; [3]; [4]; [5]. These include differences in male-female speech style during both oral [6]; [7]; [8]; and written or electronic discourse [9]. Gender related language patterns may be influenced by a host of factors such as ethnicity, class, region, and age coupled with differences in personality, professional training and sexual orientation [10]. Nevertheless, results from a variety of studies suggest that females use more emotion words [11], more hedge words such as “I think,” “kind of,” and “sort of,” more polite forms such as “would you mind if I?” or “if you don’t mind,” [12]; [13], and more tag questions such as “you don’t want to go out, do you?” or “you aren’t late, are you?” [14]. In contrast, males tend to make more frequent references to quantity, use non-standard forms or make grammatical errors, and provide opinions and justifiers [15]; [9].

Research shows that language use is affected by the nature of the task and the conversational participants [16]. As a result, the use of gender preferential language may be more common in same-gender versus mixed gender dyads [6]; [17]. For instance, Thomson and Mulachver (2001) found that during electronic conversations with a member of the other gender, both males and females changed their language style toward the gender preferential style of their conversational partner [9].

Regardless of language style, the question of who uses more words (males or females) remains open for debate. Mehl, et al., 2007 suggested that both males and females use an average of 16,000 words per day [18]. Similarly, during structured interactions, where talk has no particular goal, males and females contribute equally to discussions. However, when group status is salient and conversations are task oriented, males tend to dominate discussions, for example in a classroom [19], on an electronic mailing list [20], and during online conversations [2]. Other authors have suggested that familiarity with the discussion topic affects verbal behavior. For example, Dovidio et al (1988) reported that males spoke a greater amount of time and initiated speech more during discussions of typically masculine tasks, and females spoke longer and initiated speech more during typically feminine discussion tasks [21].

If, as some propose, females talk more than males, it would be consistent with a 2013 investigation at The University of Maryland in which researchers identified a FOXP2 gene associated with the acquisition of speech and language in humans and vocal production in birds and mammals. Results indicated a significantly higher level of FOXP2 (30% higher) in the left hemisphere of 4 year old girls than in age matched boys [22]. Consistent with these findings, Brizendine, author of *The Female Brain*, reports that women speak an average of 20,000 words per day versus 7,000 words for men [23]. However, though these numbers have been widely reported by the national media, few studies have focused on gender differences related to frequency of word use during natural communicative daily interactions or during monologues or personal narratives related to autobiographical narratives.

Research into the number of words used by males and females has been restricted by measurement systems and technologies; small language samples have been typically obtained during contrived conversational tasks, or brief observations in clinical or classroom settings [24]. Analysis of this limited data provides a snapshot, but does not provide researchers insight into the most subtle and complex of human behaviors, language. The main purpose of this study was to use the LENA, an automated speech monitoring and assessment system, to obtain and compare larger and smaller speech samples from males and females. Our larger samples were of social discourse obtained during six-hour conversational language samples in the natural environment. Shorter samples were obtained in the laboratory in order to study the narrative and expository skills of males and females across three types of language monologues: a) reminiscent monologues that required participants to use narrative skills to recount past experiences; b) story retelling monologues that required participants to use narrative skills to retell stories, and; c) favorite game or sport monologues that required participants to use expository skills to explain how to play a favorite game or sport.

Participants

Thirty-six full time college students, ages 18-28 (18 males and 18 females) who lived on campus were recruited from college classrooms at The University of Akron. No participant reported current speech, language, hearing, or other medical disorders. The project was approved by the institutional review board at The University of Akron.

Method

The research design was a nonintervention descriptive design case control study. Data was collected using LENA Digital Language Processors (DLPs) which allow for recording and analysis of language samples in natural environments and without researchers being present. The LENA recordings were transferred to a computer via a USB port and analyzed by the LENA software.

Recordings were made: 1) over 6 hours of a typical week day in the natural environment, and, 2) during three interviews in the laboratory during which participants engaged in a series of structured language tasks: a) a reminiscent task during which subjects were asked to describe the house they grew up in; b) an expository task during which subjects were asked to describe their favorite game or sport, and; c) a narrative task during which subjects were asked to retell a short story. To avoid bias, participants were unaware of the purpose of the study, and were told only that the researchers wished to record their speech.

Natural Environment Sample: Spontaneous Conversation

During the six-hour recordings, the spontaneous language conversations of subjects were recorded during a typical day. Because they “carried” (wore) LENA Digital Language Processors (DLPs), the participants are henceforth called carriers for the spontaneous condition. Data from the LENA program was downloaded into the LENA software program to produce an automatic count of the total number of words produced by each subject.

Structured Language Monologues: Narrative and Expository Skills

Narrative and expository language monologues were also recorded using the LENA. Specifically, recordings were made in the laboratory during the three monologues: a reminiscent monologue, a favorite game or sport, and a story-retelling monologue. The LENA recordings were transcribed by trained graduate students and then analyzed via the Systematic Analysis of Language Transcripts (SALT) program [25]. Utterances for each of the three monologue conditions were segmented into T-units. A T-unit is defined as one main clause and any subordinate clauses attached to it (Hunt, 1970). All maze behaviors such as false starts (e.g., um , uh) and revisions (e.g., houses, I mean, homes) were excluded from the analyses.

Reminiscent Monologue: Narrative Skills

During the reminiscent task, the examiner asked participants to recall and describe their childhood house. This task required participants to use narrative skills to recount past experiences. Task procedures were adapted from Harris (1997) [26] (see Appendix A).

Favorite Game or Sport Monologue: Expository Skills

During the expository task, the examiner asked participants to describe a favorite game or sport and talk about it in response to a series of questions. This task required participants to use expository language skills to convey factual information. Task procedures were adapted from Nippold et al (2008) (see Appendix B) [16].

Story Retelling Monologue: Narrative Skills

During the story-retelling task, the examiner read aloud a short story about a “Fast-Food Restaurant” to the participants. After listening to the story, the participants were asked to retell the story to the examiner. This task required participants to use narrative language skills to retell a story. Task procedures were adapted from Nippold et al (2009) (see Appendix C) [27].

Results

Total words spoken in 6-hour natural environment sample

A total of 305,383 words were spoken when the carriers were males and a total of 340,125 words were spoken when the carriers were females. When males were carriers, it might be expected that 50% of words recorded would have been spoken by the male carrier, 25% of the remaining words would be spoken by male communication partners and 25% would be spoken by female communication partners. This would result in 75% male spoken words (by the male carrier and male communication partners) and 25% female spoken words (by female communication partners). In fact, the data showed that when males carried the device, 83% of the words were spoken by males and only 17% of the words were spoken by females. In contrast, when females

were carriers, 76.9% of the words were spoken by females (the female carrier and female communication partners) and 23.1% were spoken by males (communication partners).

Total words spoken during three monologues: reminiscent, expository, and narrative

During descriptions of the house they grew up in, males used significantly more words than females; 669 mean words by males and 428 mean words by females. In contrast, there was no significant difference in number of words spoken by males and females when describing their favorite game/sport (507 for males and 303 for females) or retelling a story (75 for males and 65 for females).

Type Token Ratio (vocabulary diversity) produced during three monologues: reminiscent, expository, and narrative

During the descriptions of the house they grew up in, there was no significant difference in vocabulary diversity. The mean Type Token Ratio (TTR; calculated by dividing the number of different vocabulary words by the total number of words spoken), was .375 for males and .420 for females. There was no significant difference in either the mean TTR for the favorite game/sport (.39 for males and .44 for females) or the story retelling condition (.57 for males and .57 for females).

Clausal Complexity (clausal density)

During the descriptions of the house they grew up in, there was no significant difference in clausal complexity defined as the number of sentences that included complex subordinate and relative clauses. The mean clausal density measure was .359 for males and .335 for females. There was no significant difference in clausal complexity for both the favorite game/sport: .63 for males and .53 for females and the story retelling condition: .277 for males and .315 for females.

Total Number of T-Units (number of sentences)

During the description of the house they grew up in, there was a significant difference in the total number of sentences (T-units) spoken for the monologue conditions for the reminiscent task: 46.61 T-Units for females and 73.5 for males. There was no significant T-Unit difference for the other two monologues: favorite game/sport: mean of 28.38 for females and 46.44 for males), and story retelling (mean of 6.88 T-Units for females and 8.16 for males).

Mean Length of Utterance in T-Units (sentence length)

During the descriptions of the house they grew up in, there was no significant difference in sentence length. The mean of MLU-T units, defined as the mean length of T-Units in morphemes, was 9.84 for males and 10.27 for females. The favorite game/sport and story retelling monologues yielded similar non-significant results, means of 12.43 for males and 12.33 for females in the favorite game/sport and means of 10.99 for males and 10.92 for females in the story retelling condition.

Discussion

This study extends research comparing male and female talk by comparing the number of words used *during spontaneous conversations in the natural environment* when subjects were free to engage in discussions with

self-selecting conversational partners and topics. Males and females used similar numbers of words during a normal six-hour day. This data is not consistent with some of the previous research and the popular notion that females use more words than males. However, while word counts were similar, a possible gender bias was seen during spontaneous conversation. When males were carriers, 83% of the words spoken were by males. Similarly, when females were carriers, 77% of the words were spoken by females. This indicates that males spoke more to males, and females spoke more to females during a typical six-hour day, suggesting possible differences in the gender of communication partners.

This study further extends research in gender and language differences by analyzing male and female narrative and expository skills during three types of language monologues. For the story retelling, reminiscent task, and favorite game task monologues, in depth analyses of both semantic and syntactic complexity yielded similar results for males and females for measures of vocabulary diversity, clausal complexity, and the mean length of utterance of sentences. In contrast, for the reminiscent narrative monologues, males spoke significantly more words and sentences than females. This finding is consistent with Leaper and Ayres (2007) who found that males spoke more words than females during contrived tasks [24]. It is possible that males found the topic of their childhood homes more interesting than females, and in turn, produced more words and sentences for the reminiscent task.

Future research is needed to determine whether genderlect/language style for male and females during spontaneous conversation (e.g., use of emotion words, opinions, etc) is influenced by the nature of the conversational topics, partners, and settings.

References

- [1] Cohen, S.J. (2009). Gender differences in speech temporal patterns detected using lagged co-occurrence text-analysis of personal narratives. *Journal of Psycholinguistic Research*, 38, 111-127.
- [2] Guiller, J. & Durndell, A. (2006). 'I totally agree with you': Gender interactions in educational online discussion groups. *Journal of Computer Assisted Learning*, 22, 368-381.
- [3] Lakoff, R.T. (1975). *Language and women's place*. New York: NY: Harper & Row.
- [4] Mulac, A. Bradac, J.J. & Gibbons, P. (2001). Empirical support for the gender as culture hypothesis: An intercultural analysis of male/female differences. *Human Communication Research*, 27, 121-152.
- [5] Tannen, D. (2005). *Conversational style: Analyzing talk among friends*. NY: Oxford University Press.
- [6] Bilous, F. R. & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviors of same and mixed gender dyads. *Language and Communication*, 8, 183-194.
- [7] Labov, W. (1990). *Talking power: The politics of language*. New York: Basic Books.
- [8] Lakoff, R.T. (1977). Women's language. *Language and Style*, 10, 222-247.
- [9] Thomson, R. & Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology*, 40, 193-208.

- [10] Tannen, D. (1996). Researching gender-related patterns in classroom discourse. *TESOL Quarterly*, 30, 341-344.
- [11] Anderson, K.J. & Leaper, C. (1998). Emotion talk between same and mixed gender friends. *Journal of Language and Social Psychology*, 17, 419-448.
- [12] Crosby, F. & Nyquist, L. (1977). The female register: An empirical study of Lakoff's hypothesis. *Language in Society*, 6, 313-322.
- [13] Macaulay, M. (2001). Tough talk: Indirectness and gender in requests for information. *Journal of Pragmatics*, 33, 293-316.
- [14] Cameron, D., McAlinden, F. & O'Leary, K. (1989). Lakoff in context: The social and linguistic functions of tag questions. In J. Coates & D. Cameron (Eds.), *Women in their speech communities: New perspectives on language and sex*. London, UK: London.
- [15] Mulac, A., Siebold, D. & Farris, J. (2000). Female and male managers' and professionals' criticism giving differences in language use and effects. *Journal of Language and Social Psychology*.
- [16] Nippold, M.A., Mansfield, T.C., Billow, J.L. & Tomblin, J.B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech Language Pathology*, 17, 356-366.
- [17] itzpatrick, M.A., Mulac, A. & Dindia, K. (1995). Gender-preferential language use in spouse and stranger interaction. *Journal of Language and Social Psychology*, 14, 216-222.
- [18] Mehl, M.R., Vazire, S., Ramfrez-Esparaza, N., Slatcher, R.B. & Pennebaker, J.W. (2007). Are women really more talkative than men? *Science*, 317, 82.
- [19] Galvin, S.M., Dolly, M.R. & Pula, J.J. (2013). Genderlect and participation in the college English classroom. *The Delta Kappa Gamma Bulletin*, 79, 22-30.
- [20] Herring, S.C. (2000). Gender differences in Computer mediated communication (CMC): Findings and implications. *Computer Professionals for Social Responsibility*, 18.
- [21] Dovidio, J.F., Heltman, K., Brown, C.E. & Ellyson, S.L. (1988). Power displays between women and men in discussions of gender linked tasks: A Multichannel study. *Journal of Personality and Social Psychology*, 55, 580-587.
- [22] Bowers, J.M, Perez-Pouchoulen, M., Edwards, N.S. & McCarthy, M.M. (2013). Foxp2 mediates sex differences in ultrasonic vocalization by rat pups and directs order of maternal retrieval. *The Journal of Neuroscience*, 20, 3276-3283.
- [23] Brizendine, L. (2006). *The female brain*. NY: Broadway Books.
- [24] 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*.

[25] Miller, J. & Chapman.R. (2003). *Systematic Analysis of Language Transcripts (SALT)*, Madison, WI.

[26] Harris, J.L. (1997). Reminiscence: A culturally and developmentally appropriate language intervention for older adults. *American Journal of Speech Language Pathology*, 6, 19-25.

[27] Nippold, M.A., Mansfield, T.C., Billow, J.L. & Tomblin, J.B. (2009). Syntactic development in adolescents with a history of language impairments: A follow-up investigation. *Journal of Speech Language Pathology and Audiology*, 18, 241-251.

Appendices

Appendix A: Reminiscent Task Instructions for Data Collection (Adapted from Harris, 1997)

“Visualize and describe the house you grew up in. Take us on an imaginary walk through.” (hit record)

They answer, then pause:

“Tell me about the people in your household such as your siblings, grandparents, and others. Did you have a room of your own? Who shared your room?” (hit record)

They answer, then pause:

“Who were your neighbors? How far from your house was your school?” (hit record)

They answer, then pause:

“REMEMBER THAT WE ARE TALKING ABOUT THE HOUSE YOU GREW UP IN”

“Tell me about where you bought your groceries.” (hit record)

They answer, then pause:

“What did you do for recreation and entertainment?” (hit record)

They answer, then pause:

“What kinds of chores were assigned to you? And What did you like best about your childhood home?” (hit record)

Appendix B: Favorite Game or Sport Instructions for Data Collection (Adapted from Nippold et al, 2008)

“I am hoping to learn what people know about certain topics. There are no penalties for incorrect answers.

“What is your favorite game or sport? (hit record)

They answer, then pause:

“Why is _____ your favorite game/sport?” (hit record)

They answer, then pause:

“I’m not too familiar with the game of _____, so I would like you to tell me all about it. For example, tell me what the goals are, and how many people may play a game. Also, tell me about the rules that players need to follow. Tell me everything you can about the game of _____ so that someone who has never played before will know how to play.” (hit record)

They answer, then pause:

“Now I would like you to tell me what a player SHOULD do in order to WIN the game of _____. In other words, what are some key strategies that every good player should know?” (hit record)

They answer, then pause.

Appendix C: Story Retelling Instructions for Data Collection (Adapted from the peer conflict task (Nippold, Mansfield , Billow & Tomlin, 2007)

“People are always running into problems with others at school, at work, and at home. Everyone has to work out ways to solve problems. I am going to read a story to you that illustrates these types of problems. I would like you to listen carefully, and be ready to tell the story back to me, in your own words.”

(Male Story Version)

“Mike and Peter work at a fast-food restaurant together. It is Mike’s turn to work on the grill, which he really likes to do, and it is Peter’s turn to do the garbage. Peter says his arm is sore, and asks Mike to switch jobs with him, but Mike doesn’t want to lose his chance on the grill.” (hit record)

They answer, then pause:

“Now I’d like you to tell the story back to me in your own words. Try to tell me everything you can remember about the story.” (hit record)

(Female Story Version)

“Jane and Kathy work at a fast food restaurant together. It is Jane’s turn to work on the grill, which she really likes to do, and it is Kathy’s turn to do the garbage. Kathy says her arm is sore, and asks Jane to switch jobs with her, but Jane doesn’t want to lose her chance on the grill.” (hit record)

They answer, then pause:

“Now I’d like you to tell the story back to me in your own words. Try to tell me everything you can remember about the story.” (hit record)