CHAPTER XX

Socio-linguistic factors and gender mapping across real and virtual world cultures

Aaron Lawson[1], Kyle Leveque[1], John Murray[1], Wen Wang[1], Nick Taylor[2], Jennifer Jenson[2], Suzanne de Castell[3]

[1]SRI International, Menlo Park, CA USA
[2]York University, Toronto, ON, Canada
[3]Simon Frasier University, Vancouver, BC, Canada
{aaron.lawson, kyle.leveque, john.murray, wen.wang}@sri.com;
nicktaylor@gmail.com, jjenson@edu.yorku.ca, decaste@sfu.ca

ABSTRACT

This study examines a large corpus of online gaming chat and avatar names to explore gender differences in virtual world (VW) language use. In particular, we examine the relevance of socio-linguistic observations of gender in face-to-face conversation to the contemporary space of VW chat interactions in online gaming environments. In addition, we study the relationship between a player's gender and naming decisions for online avatars in terms of linguistic observations based on sound symbolism. Analysis shows that many of the existing socio-linguistic claims about gender and speech also hold true in the VW for many of the categories posited (e.g. swearing, hedging, empathy). For avatar naming conventions, results showed that these rules could predict gender with a high average accuracy (>0.7) for both males and females. Applying the same rules to avatar names from individuals whose real world (RW) and VW gender were different still enabled detection of RW gender at a similar high rate of accuracy, despite mismatched gender. We conclude that the predictions of socio-linguists about gender-linked behaviors and decisions in RW conversational interactions largely transfer across subcultures to VW environments such as online gaming chat and avatar naming conventions.

Keywords: gender and speech, virtual world culture, socio-linguistics

1 INTRODUCTION

This study examines a large corpus of online gaming chat and avatar names to explore gender differences in virtual world (VW) language use. In particular, we examine the relevance of traditional socio-linguistic observations of males and females in face-to-face conversation to the contemporary space of VW chat interactions in online gaming and collaborative environments. In addition, we study the relationship between a player's real world (RW) gender and naming decisions for online personas, or avatars, in the light of linguistic observations based on sound symbolism and naming conventions. The approach taken in this study focused on applying socio-linguistic claims and observations to develop discourse features for characterizing gender in virtual world chat and looking at other linguistic factors, such as choice of avatar name, to detect gender trends. To expedite the development of features we examine the rich empirical claims of the socio-linguistic literature to identify known factors that have tended to correlate with male or female speech. A primary goal of this study is to determine whether these findings apply in the physically distant universe of VW interactions.

1.1 VERUS Project

This study (Table 1) is situated within the larger context of the VERUS project (Dieterle 2011). The Virtual Environment Real User Study (VERUS) is a research project conducted in collaboration with SRI International, Simon Fraser University, York University, and Nottingham University Business School. The goal is to understand what can be learned about individuals and groups when observing their activities and behaviors in elaborate multiplayer online games and other virtual world environments. Which in-game features are most predictive of real-world characteristics? How might these technologies be utilized effectively for training or learning environments? How do these features and models perform when used on numerous games?

Table 1 Data distribution of virtual world chat in the VERUS project.

Game	Turns	Talkers	Tokens
Guardian Academy	914	57	2688
Sherwood	13,149	271	57,843
SecondLife	79	4	392
WoW	2337	117	56,036
Total	**11214**	**445**	**89,521**

Initial studies have included volunteers playing different games, including Guardian Academy, Second Life, Sherwood, and World of Warcraft. In-game features include communication styles, movement patterns, engagement behavior, and avatar

selection. Data for this study was taken from the VERUS-internal corpus of virtual world chat and avatar demographics information from the Sherwood and Guardian Academy worlds only. Avatar names used in this study were chosen by participants when setting up their character information only for games on the VERUS server environment.

2 THEORETICAL BACKGROUND

2.1 Gender and Discourse Factors

There is a long history of language and gender studies in the field of socio-linguistics. This is widely credited as having begun with Robin Lakoff (1975) in the 1970's and continues today with research by Deborah Tannen (1984, 1994), Deborah Cameron (2006), and specifically in computer-mediated language use by Susan Herring (1994, 2006). Of particular interest for this study, this research makes empirical claims about the differences between male and female speech at the lexical and discourse levels and provides a testable starting point for features that can be used to distinguish male versus female participants in virtual world chat.

Early research focused on observing the role of gender in speech as a way of characterizing men's versus women's behavior, be it innate or learned. Later work has come to see many of these linguistic differences in speech as manifestations of power, status or role differences, rather than as purely gender-based ones. An important early study on language and gender was Lakoff's *Language and Women's Place* (1975). This work lays out specific features of women's speech as observed by Lakoff in her analysis of conversation and provides several empirical claims about gender and use of language. Chief among those characteristics associated with female speech are

- Hedging, hesitation, greater uncertainty and expressions of uncertainty
- Greater use of polite forms and expressions associated with politeness
- More use of question forms and intonational prosody
- Frequent use of apologies, even in cases when no fault could be found
- Greater use of modal verbs (can, could, would, should).
- Avoidance of insults or cursing

Lakoff's work points out general conversational strategies in women's speech, such as indirectness, avoiding confrontation, and avoiding absolute statements that underlie these observations.

O'Barr and Atkins (1980) examined language use in courtroom settings and specifically investigated Lakoff's conclusions about those characteristics identified as "women's language". They found that the evidence did not support a taxonomy of traits that could be unequivocally associated with women. Rather, they found that these traits were not confined to women, but more reflected power differences and status in a given situation, rather than gender. Likewise, Tannen, a student of

Lakoff who follows a similar methodology in her work (Tannen 1984, 1994) came to slightly different conclusions from Lakoff. For Tannen, the factors that Lakoff generally associated with gender are best described in terms of dominance in conversation, and not necessarily gender *per se*, although they may manifest themselves along gendered lines in conversations where males tend to be dominant.

Keith and Shuttleworth summarize traits more common to women and traits more common to men in *Living Language* (2000). They associate women's speech with more talk (this differs from Lakoff), polite language, asking of questions, providing supportive statements and expressions of empathy. In male language use they noted more swears, more insults, and the tendency to give commands. Susan Herring, who works more specifically in the area of online discourse, chat and bulletin boards (1994, 2006), characterizes male speech in these arenas as revolving around adversariality and women's speech as being associated with attenuation.

Male on-line discussions

- Put-downs
- Strong, often contentious assertions
- Lengthy and/or frequent online postings
- Self-promotion and sarcasm
- Name calling and personal insults
- Challenging an interlocutor's "face"

Female online discussions

- Hedging
- Apologizing
- Asking questions rather than making assertions
- Thanking
- Agreeing

The work of Herring demonstrates that many of the features first identified by Lakoff are also at play in online communications that do not involve direct face-to-face interactions. Based on this latest research, this study hypothesizes that some of those features identified by researchers as indicative of gender or social status in conversational speech will also be applicable to online discourse in the virtual world. Further this demonstrates that we may be able to leverage these traits as features to help automatically distinguish male from female participants in virtual world environments.

2.2 Sound Symbolism and Gender

Sound symbolism refers to the relationship between linguistic units of sound, either on the phonemic level, feature level or acoustic trait, and a extra-linguistic quality, such as gender or size. Some of these relationships are clearly iconic, as in the case of onomatopoeia; others are less obvious, such as the common relationship between high, front vowels (/i/, for example) and words dealing with small size or female gender. This second relationship is hypothesized to revolve around the high

frequency of the second formant in high front vowels, its perception as being overall higher in frequency, and the association of children's and women's voices with higher relative frequency (Gordon & Heath, 1998). The association between low and back-rounded vowels and large size or masculinity across languages is based on similar reasoning–the reduction in frequency of the second formant due to lip rounding or lowering of the tongue gives the perception of a lower frequency sound, associated with larger size or male gender. Sibilant and strident consonants, such as /sh/, are characterized by very high frequency noise due to small and focused frication.

The field of sound symbolism has a long history in the study of language. Jespersen (1922) provides a summary of the earliest research and theories, while Ohala et al. (1994) present late 20th century views on the phenomenon. Current debates on sound symbolism deal with the notion of whether most sound symbolic trends are due to physical, biological or 'innate' factors or whether they are a function of cultural promulgation. For this study the core difference is not important and is probably based on a combination of acoustic, auditory, articulatory, and cultural factors.

Gender and sound has been the focus of significant research. Jakobson (1990) looked at the origin of sound symbolism in child language learning and in the early association of mothers with nursing, and hence the frequency of maternal terms cross-linguistically containing the bilabial nasal /m/, the sound made by babies when nursing, and its association with enjoyable food, as in the expression "mmm". Romaine (1999) details findings about the relationship between gender and certain vowels and consonants, in particular noting that both the manner and place of articulation are relevant, not in themselves but in the way they impact the acoustic perception of the sound. She cites back vowels and back consonants as being associated with masculinity and labial consonants as being associated with femininity, among other trends.

3 RESEARCH QUESTIONS

Our research questions can be broken down into three discrete queries: 1) do the empirical claims of traditional socio-linguistic literature about gendered speech hold true in the VW and can they be used to identify the real world (RW) gender of an individual? 2) Can we identify the RW gender of an online individual based on that individual's choice of avatar names? 3) If so, do the conventions and decisions people make for online avatars still hold true when the VW gender and the RW gender are different (RW male plays female in VW or vice versa)?

4 APPROACH

In the first part of the study, a set of ten lexical-level and discourse-level features was developed based on gender claims from the socio-linguistic

literature. These included swearing, insults, slurs, modal verbs, apologizing, expressions of uncertainty and empathy, and questioning. These features were determined based on how prominent they were in the research and how testable they might be. Each turn in the VW chat database was evaluated for these features to determine the probability that the turn came from a male or a female. Since the frequency of males is higher than females in this data set (55% to 45%) results are presented using the adjusted probability for each gender. The features are listed in Figure 1.

For the investigation of the relationship between gender and avatar naming, thirteen rules, largely based on observations from the sound symbolism research detailed above, were developed-four for females and nine for males. These included phonetic rules such as female names ending in low vowels, male names ending in back vowels, male names containing 'z' and 'x', and female names containing 'sh'. In addition, we included more basic rules, such as the use of female names for female players and male names for male players, based on 2010 U.S. census data. These rules are listed in Table 2.

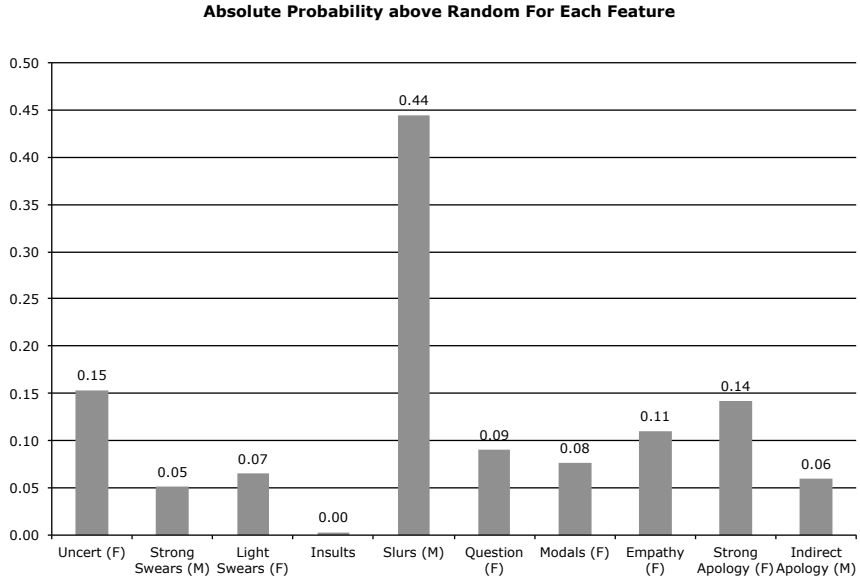

Figure 1 Contribution of each feature to identifying male (M) and female (F) in virtual world chat

5 RESULTS AND ANALYSIS

Analysis of the results showed that many of the existing socio-linguistic claims about gender and discourse also hold true in the VW: women were much more likely to use modal verbs, ask questions, use expressions of uncertainty, and use strong apologies than were males. Males were much more

likely to use strong swears, slurs, and indirect apologies. The results also demonstrate that some of the categories suggested in the socio-linguistic studies may be too coarse. For example, women are claimed to apologize more frequently than men. However, if one analyzes the types of apologies that occur it become clear that direct apologies (e.g. "I'm sorry") are more typical of female players, while indirect apologies ("oooops!" or "my bad!") are more associated with males. Similarly, with swears there was a breakdown between the types of words used: light swears were associated more with females and strong swears more with males. This observation is actually in keeping with the observations that men are more comfortable with profanity than females, as many of the 'light swears' represent approaches to avoid offensive cursing. Slurs were the category most strongly associated with males, and most slurs were homophobic in nature.

In the second part of this study, the relationship between RW gender and avatar names was examined. For those avatars whose gender in the real world and virtual world were the same, results showed that these rules could predict gender with a high average accuracy (> 0.7) for both males and females.

Table 2 Characteristics of male versus female avatar name choice

Characteristic	**Gender**	**Precision**
Ends with fricative consonant	M	0.9
Is a female US 2010 census name	F	0.89
Ends in "a"	F	0.85
Contains a title of nobility	M	0.85
Is a male US 2010 census name	M	0.85
Ends in "er"	M	0.82
Ends in back vowel	M	0.81
Contains 'x' or 'z'	M	0.79
Ends in back or alveolar stop	M	0.75
Ends in any consonant	M	0.68
Ends in "y"	F	0.66
Begins with capital	M	0.62
Contains palatal fricative (sh)	F	0.61

The highest precision sound-based rules deal with word endings, with those words ending in a fricative consonant being strongly male, and words ending in the central vowel schwa (represented orthographically with "a") being strongly female. Applying the same rules to avatar names from individuals whose RW and VW gender were different still enabled detection of RW gender at a similar high rate of accuracy (>.7), despite mismatched gender. This result was surprising and may show that avatar gender was not playing a significant role in the

players' online persona.

6 CONCLUSIONS

The major finding of this study is that the predictions of socio-linguistics about gender-linked behaviors and decisions in RW conversational interactions largely transfer across subcultures to the VW environments explored in this study. In the area of online gaming chat many of the gender-based characteristics observed by Lakoff, Tannen, Herring, and others, broadly grouped as 'attentuative' and 'accommodating' as opposed to 'adversarial' or 'aggressive', were evident in this data. This was true for specific grammatical features such as the greater frequency of use of modal verbs for women and for lexical choices such as the greater frequency of strong swears for men. In the case of avatar naming conventions, many trends in sound-gender relationship were manifest in the choice of avatar naming. It is unclear whether these trends are linguistic "universals", or if they are reflections of deeply engrained cultural conventions or a function of traits specific to Indo-European languages. Regardless of their ultimate origin, sound-symbolic trends were instrumental in determining the RW gender of the participants in this study based on their choice of avatar name. This finding held both for those whose RW gender and VW gender were the same and those for whom the VW and RW gender differed.

Future work in this area is expected to include an exploration of how avatar names can link real-world individuals across multiple virtual world personas and how avatar naming conventions pattern across different age groups. Further, the authors intend to examine the relationship between age group and linguistic factors in chat: lexicon choice, typography, pronominal reference, and the acquisition of literacy. Certainly, the rich data resources available in online environments present a new and evolving domain for the study of cross-cultural communication and the understanding of individual differences in decision making between real-world and virtual-world environments.

ACKNOWLEDGMENTS

The authors acknowledge the Air Force Research Laboratory at Wright Patterson Air Force Base for support of this research.

REFERENCES

Cameron, D. 2006. "Theorizing the female voice in public contexts", in *Speaking Out: The Female Voice in Public Contexts*, ed. Judith Baxter. Houndmills: Palgrave.

Dieterle, E. and Murray, J. 2011. "Virtual Environment Real User Study: Design and Methodological Considerations and Implications", Journal of Applied Learning Technology, Vol. 1, No. 1, 19-25.

Gordon, M. and Heath, J. 1998. “Sex, Sound Symbolism, and Sociolinguistics”, Current Anthropology, Vol. 39, No. 4, August/October.

Herring, S. C., and Paolillo, J. C. 2006. “Gender and genre variation in weblogs”. *Journal of Sociolinguistics*, 10(4), 439-459.

Herring, S. 1994. “Gender Differences In Computer-Mediated Communication: Bringing Familiar Baggage To The New Frontier”, American Library Association Annual Convention, Miami, FL.

Jakobson, R. 1990. “Why Mama and Papa?” in *On Language*, L. Waugh and M. Monville-Burston, Eds.

Jespersen, O. 1922. *Language: Its Nature, Development and Origin*. London: Allen and Unwin.

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

Ohala, J., Hinton, L. and Nichols, J. 1994. *Sound Symbolism*. New York: Cambridge University Press.

O’Barr, W. M., and Atkins, B. K. (1980). “Women’s language” or “powerless language”? In S. McConnell-Ginet, N. Borker, & R. Thurman (eds.), *Women and Language in Literature and Society,* New York: Praeger, 93-110.

Romaine, S. 1999. *Communicating Gender*. London: Lawrence Erlbaum Associates.

Shuttleworth, John and Keith, George. 2000. *Living Language*. Hodder Education.

Tannen, D. 1994. *Gender and Discourse.* NY & Oxford: Oxford University Press.

Tannen, D. 1984. *Conversational Style: Analyzing Talk Among Friends*. Norwood, NJ: Ablex.