

UNIVERSITY OF HELSINKI

Linguistic Assimilation to Gendered Avatars

Designing A Case Study in World of
Warcraft



Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistinen		Laitos – Institution – Department Nykykielten laitos	
Tekijä – Författare – Author Susi Vaasjoki			
Työn nimi – Arbetets titel – Title Linguistic Assimilation to Gendered Avatars: Designing A Case Study in World of Warcraft			
Oppiaine – Läroämne – Subject Englantilainen filologia			
Työn laji – Arbetets art – Level Pro Gradu		Aika – Datum – Month and year Kesäkuu 2016	Sivumäärä– Sidoantal – Number of pages 55
Tiivistelmä – Referat – Abstract			
<p>Itsekategorisoinnin teorian mukaan yksilöt sopeuttavat käyttäytymistään siihen, miten olettavat ulkopuolisen havainnoitsijan näkevän itsensä. Deindividuaation teorian mukaan, jos yksilöiden olemusta manipuloidaan esimerkiksi univormuilla, rooliaisuilla tai videopelihahmoilla, yksilöt korjaavat käytöstään manipulaatiota myötäillen. Efektia kutsutaan assimilaatioksi, ja sen on havaittu koskevan useita käyttäytymisen alueita, kuten epäsosiaalista ja prososiaalista käytöstä, itsevarmuutta, läheisyyden kokemista, sukupuolettamista mutta myös asenteita, kuten seksististen ja rasististen ennakkoluulojen hyväksyttävyyttä. Assimilaatiohypoteesin mukaan videopelaajat omaksuvat niitä tapoja ja asenteita, jotka olettavat muiden kokevan pelihahmolleen sopivaksi.</p> <p>Tässä tutkimuksessa selvitettiin assimilaatiohypoteesin toteutumista luonnollisessa videopeliympäristössä, jossa pelaajat pelasivat joko oman sukupuolensa mukaista tai siitä poikkeavaa hahmoa. Tutkittavana muuttujana oli pelaajien sukupuolittunut kielenkäyttö ja sen mahdollinen muutos pelihahmon sukupuolen mukaan. Sukupuolta käsitellään tutkimuksessa stereotyyppisen binäärioletuksen mukaisesti, koska itsekategorisointi ja siitä nouseva assimilaatio perustuvat nimenomaan stereotyyppisiin oletuksiin, eivät empiiriseen dataan. Vastaavasti lingvistiset mittarit on valittu folklingvististen stereotyyppien kvantifiointiksi, ei empiirisen datan perusteella. Naistyyppillisen puheen malli perustuu aikaisemmassa tutkimuksessa käytettyyn Lakoffin naisen kieltä noudattavaan malliin. Sen komplementiksi on tässä tutkimuksessa rakennettu hegemonisen maskuliinisuuden esittämiseen liittyvien piirteiden malli, joka perustuu pääosin Coatesin ja McEneryn havaintoihin miestyypillisenä pidetystä puheesta.</p> <p>Tutkimusaineistona oli pieni kokoelma vapaaehtoisten World of Warcraft -pelaajien keskustelulokeja (n = 6 vastaajaa, n = 8 lokia). Lokien lisäksi pelaajilta pyydettiin suppeat henkilötiedot (mm. ikä, sukupuoli, kansallisuus, hahmon nimi ja palvelin Warcraft Armory -hakuja varten). Naistyyppillisten piirteiden mittaamiseen käytettiin aikaisemmassa tutkimuksessa vakiintunutta lakoffilaista kolmen piirteiden mittaria (anteeksipyynnöt, tunnelmaisuus sekä pehmentävä/epävarmuutta ilmaiseva kielenkäyttö). Miestyyppiselle kielelle laadittiin kolmipiirteinen kokeellinen mittari (huumori, kilvoittelu, kirosuhteet). Piirteet koodattiin lokiteksteistä manuaalisesti, minkä jälkeen assimilaatiohypoteesin toteutumista testattiin tilastollisen t-testin avulla. Pieni aineisto ja lokien sisällöllisistä rajoitteista (mm. tietoista esittämistä sisältävien roolipelitilanteiden ulosulkeminen) johtuva datakato tarkoittavat, että aineiston tilastollinen näyttövoima on kuitenkin heikko, ja tulokset soveltuvat lähinnä menetelmän koeajoon.</p> <p>Assimilaatiohypoteesi toteutui aineistossa vain rajoitetusti. Raakatassa pehmentävä kielenkäyttö ja huumori vaikuttivat noudattavan hypoteesia, mutta tulokset eivät olleet tilastollisesti merkittäviä. Tunnelmaisuus ja kilpailullinen kielenkäyttö toimivat tilastollisesti merkittävällä tasolla assimilaatiohypoteesin vastaisesti. Lokien lähiluku antaa syytä olettaa, että tämä johtuu polarisointuneesta näytteestä: kilpailullinen pelitilanne sekä pelaajan oma voimakas sitoutuminen kilpailullisiin tavoitteisiin korreloivat tilastollisesti merkittävällä tasolla kilpailullisen kielenkäytön kanssa, hahmon tai pelaajan sukupuolesta riippumatta. Kilpailullinen pelitilanne vaikuttaa myös vähentävän pelaajan tunnelmaisuutta (mikä sinänsä sopii assimilaatiohypoteesiin, jos kilpailullinen pelaaminen mielletään maskuliiniseksi toiminnaksi). Anteeksipyynnöiden ja kirosuhteiden data ei riittänyt analyysiin.</p> <p>Luotettavampien tulosten saamiseksi menetelmää tulisi kehittää varmistamalla jokaiselta vastaajalta näyte sekä saman- että erisukupuolisella hahmolla, jolloin kukin pelaaja toimisi omana kontrollinaan. Näytteiden määrän tulisi olla suurempi ja pelikontekstin paremmin standardoitu. Lisäksi maskuliinisen kielenkäytön mittari tulisi testata laajemmalla ja standardisoidummalla aineistolla sen validiteetin toteamiseksi. Näin toteutettuna tutkimus voisi valottaa tapaa, jolla virtuaaliset ympäristöt ja niissä toimiminen vaikuttavat käyttäjiin. Jos assimilaatiohypoteesi voidaan osoittaa todeksi, se asettaa videopelien ja virtuaaliympäristöjen käyttäjiin kohdistamista vaikutteista eettisiä kysymyksiä, jotka jokaisen kehittäjän tulisi huomioida.</p>			
Avainsanat – Nyckelord – Keywords Gendered language, World of Warcraft, assimilation, deindividuation, self-categorisation, avatar, video game, competitive language, emotional language, gameplay context			
Säilytyspaikka – Förvaringställe – Where deposited E-thesis			
Muita tietoja – Övriga uppgifter – Additional information			

Susi Vaasjoki
Pro Gradu
English Philology
Department of Modern Languages
University of Helsinki
May 2016

Index

1 Introduction.....	1
2 Theoretical background.....	4
2.1 Assimilation as a cognitive process.....	4
2.1.1 Deindividuation and self-categorisation.....	4
2.1.2 Examining assimilation through language.....	6
2.2 Language and gender.....	7
2.2.1 Female-typical language.....	7
2.2.2 Male-typical language.....	10
2.3 Corpus research into language and gender.....	13
3 Methods.....	15
3.1 Research questions.....	15
3.2 The case environment.....	15
3.3 Research ethics.....	18
3.4 Data collection and sampling.....	19
3.5 Operationalised metrics and coding.....	21
3.6 Corpus based analysis.....	22
3.7 Statistical testing.....	23
4 Data and findings.....	23
4.1 Data.....	24
4.1.1 Logged speech contexts.....	24
4.1.2 Sample length.....	30
4.2 Findings.....	31
4.2.1 Verbosity.....	32
4.2.2 Female-typical language markers.....	34
4.2.3 Male-typical language markers.....	36

4.2.4	Gameplay context as a shaper of language.....	37
4.2.5	Gendered use of parts of speech.....	41
5	Discussion.....	42
5.1	Refining the study design.....	42
5.1.1	Improving sample stability.....	43
5.2	Refining the analysis.....	45
5.2.1	Coding ambiguities for future consideration.....	45
5.2.2	Testing male-typical metrics.....	47
5.3	Implications.....	48
6	Conclusions.....	50
7	References.....	51
7.1	Bibliography.....	51
7.2	Secondary Literature.....	54
7.3	Software.....	54

Index of Images

Image 1: Concept art of player races in World of Warcraft	16
Image 2: Default user interface in World of Warcraft	17

Index of Tables

Table 1: Summary of responders and demographic data	20
Table 2: Summary of correlations and significance testing	32
Table 3: Average line-length	33
Table 4: Female-typical language	34
Table 5: Male-typical language	36
Table 6: Language change in paired samples	38
Table 7: Gameplay context	40

1 Introduction

Digital games are well on their way to becoming as integral to modern entertainment culture as television and films. The most recent study of the Interactive Software Federation of Europe showed that 48% of all Europeans and 60% of the Finnish population have at least tried video games, and that an average of 25% play some form of digital game every week. Modern video gaming is a fundamentally social pursuit: the same study states that on average 81% of European gamers play together with others, and approximately half of this group state that given the option, they will play with others rather than alone (ISFE, 2012). Sometimes, this will mean no more than flipping open a mobile phone app to play a round of Scrabble against a human opponent – but often games will be more elaborate, set in digital worlds where the player moves with relative freedom, represented by a 3D-modelled humanoid avatar. The most complex in the current crop of virtual worlds, shared by thousands of players, are referred to as MMOs, massively multiplayer online games.

MMO games create a unique language use context. Represented by graphic avatars and a self-selected screen name, players need not give away any more of their offline selves than they wish. For many, this is an opportunity to let slip the boundaries of conventional identity and play, making friends and building communities without the usual boundaries of geography, age or nationality. Many also experiment with gender, electing to play a character whose sex doesn't match their own. The practice, commonly called gender-swapping, is more popular among women than men, by a significant margin; Hussain & Griffiths (2008) cite the numbers as 54% and 68% for men and women respectively, while Lou & al. (2013) state 37% and 45,3%. Players report a variety of motives for gender-swapping: some reasons are utilitarian, like avoiding harassment or being treated better by male players (interestingly, this is reported by both men and women), but others list experimentation and plain fun (Hussain & Griffiths 2008). In environments where all bodies are pixel perfect and text chat is the primary means of interaction, gender presentation becomes largely a matter of choice.

This ease of manipulating identity makes virtual worlds a fascinating arena for research. When we change the bodies that represent us, do we also change the selves

that we use them to represent? The player's "real life" identity does carry over in some ways – studies have found that players are more likely to have friends who are of the same sex as they are, regardless of what avatars they play (Lou & al 2013), and even in games with an international player base, to play with people who live fairly close to them geographically, even in the same city (Williams, Yee & Caplan 2008). But other evidence also suggests that changes of avatar do come with matching changes of behaviour. Some of these are deliberate. In the MMO *Fairyland Online*, female characters are reputed to have better success in the game's barter-based market, which may encourage male players to create female avatars to improve their success at trade – likewise, male characters are reputed to be stronger combatants, and this may encourage female players to create male avatars if their interest is primarily in competitive gameplay (Lou & al 2013). Some male players will also create female avatars with the express intent of exploiting male courtesy and soliciting gifts (Hussain & Griffiths 2008). These behaviours are obviously designed to exploit gendered role expectations – but are there other ways in which avatars imprint on behaviour, regardless of player intent?

Cognitive psychology defines sense of self as an assessment of ourselves, evaluated as if through the eyes of a hypothetical other. If the self being presented is altered, perception of self as the member of a group should lead an individual to adopt behaviours appropriate to the in-group, and eschew the behaviours associated with the undesired out-group (Bem 1972). In conventional experiments this alteration has been achieved by means of uniforms (Frank & Gilovich 1988; Johnson & Downing 1979). The process is called assimilation.

The assimilation effect persists past the boundaries of the material world, following us into more ephemeral computer-generated spaces. The virtual names, faces and bodies we use to represent ourselves in computer-mediated interactions are assimilated as part of our perceived self, and serve to shape our behaviour much like their real-world counterparts would. Nor is the effect confined to the time we spend with an adopted virtual body: cognitive studies have produced evidence to suggest that assimilated behaviours and perceptions persist past leaving the virtual world, influencing things such as confidence and risk-taking behaviour (Yee 2007), racist ideation (Peck, Seinfeld, Aglioti & Slater 2013), self-image and rape myth

acceptance (Fox, Bailenson & Tricase 2013) and even the linguistic strategies we use to express ourselves (Palomares & Lee 2009). As our perceptions of ourselves change, so does the role we expect to be seen in – and we alter our behaviours accordingly.

The majority of research on avatar assimilation has been conducted in a laboratory, using carefully crafted test scenarios and behavioural evidence. This study examines the possibility of observing the effect in the field, using gameplay chat generated in the course of online video gaming. The paper outlines some previous research on assimilation and gendered language and makes a case for using folklinguistic beliefs on gendered language to track players' assimilation to gendered avatars. Based on this, I propose a method for evaluating assimilation to gendered avatars via a set of female-typical metrics established in existing research, and on a proposed set of male-typical metrics compiled for this study. To test the viability of the method, I examine a small set of linguistic data from a natural language setting – though due to a faulty sampling method, the final sample ($n = 8$) is too small to be considered in any way representative. Finally, the method from sampling to analysis is re-examined for possible sources of error in set-up, metrics and data coding, and refined with an eye on undertaking a more representative data set. Actual testing of the assimilation hypothesis is beyond of the scope of the study, though some provisional observations can be made.

World of Warcraft (from here on abbreviated as *WoW*) was selected as a case environment because of its overwhelming popularity: with a commercial history beginning in early 2005, over 12 million players at the height of its success and 5.5 million at present (Statista 2015), it is easily justified as one of the most persistent and popular virtual worlds. Gender is selected as an assimilation parameter because unlike other possible parameters (e.g. avatar race, which changes a character's game mechanics, or height, which impacts the size of the target an avatar presents to hostile players), gender conveys no direct tactical benefit in the game, and is more likely to be chosen by preference. Previous research in both avatar assimilation and in gender and language also offers some established metrics useful for evaluating assimilation or the lack thereof (Palomares & Lee 2009).

For comparability with previous research, gender will here be examined as a restrictive binary system (male/female). While the reality of both sex (biological) and gender (social) is more complex, *WoW* models its avatars as strictly and dramatically dimorphic – participation in the game requires the player's implicit consent to binary gendering. Unless otherwise specified, this paper will discuss gender in this limited, binary sense. The problems inherent in doing so are recognised; and though examining avatar assimilation under conditions where no matched gender option is available is an interesting premise in its own right, it falls outside the scope of this paper.

2 Theoretical background

2.1 Assimilation as a cognitive process

2.1.1 Deindividuation and self-categorisation

Cognitive theories on deindividuation predict that subjects placed in an environment that strips them of external identity cues will both overemphasise the cues that remain, and remodel their behaviour to conform to the behaviour they feel is expected of them (Bem 1972). The effect is well exemplified by the cultural practice of wearing uniforms: controlled laboratory experiments have shown that donning a uniform not only signals group membership to others, but also prompts the individual to shift their behaviour in accordance with the expectations primed by the uniform. In one experiment, subjects given a black uniform were perceived as much more aggressive and threatening, both by themselves and others, than subjects given a white uniform (Frank & Gilovich 1988). Similarly, test subjects given Ku Klux Klan-style outfits were willing to administer much longer and more severe electrical shocks than subjects given nurse uniforms, who conversely were reluctant to administer shocks at all (Johnson & Downing 1979).

Similar deindividuation and altered self-perception has been demonstrated in virtual environments, where the subject assimilates traits of their avatar's appearance and modifies their behaviour as if these were aspects of their realworld physical body. Parallel to the uniform study, subjects in a virtual environment given black-robed avatars committed significantly more antisocial acts than their counterparts whose

avatars wore white (Merola, Penas, & Hancock, 2006). Nick Yee describes a set of laboratory studies where avatar appearance is successfully used to manipulate the users' behaviour. In the first study, users given a more attractive avatar were more confident, as measured by the quantity of personal data they were willing to share, and how close they were willing to stand to their conversation partner (Yee 2007: 36-37). In the second study, users given a taller avatar showed greater assertiveness, as measured by the likelihood of their offer in an Ultimatum Game task being an unfair one (2007: 42). Following up on these studies, Yee presents a case in which the assimilated traits persist outside the virtual environment: users assigned a tall avatar for the Ultimatum Game task continued to negotiate more aggressively face to face (2007:46). If time spent in virtual bodies imprints behavioural legacies on the player, Yee argues, it is worth considering what those bodies are, and what behaviours they enforce. While embodying a tall and attractive avatar may help those with confidence issues, similar transference of antisocial behaviours, for example, would be less than desirable (2007:103).

A notable difference between Yee's work and classical deindividuation studies is that Yee examines only shifts of degree, rather than shifts of category. Tall/short and attractive/unattractive are attributes commonly understood as scales, where all subjects will fall somewhere between the extremes – Yee's experiment manipulates the degree to which the attribute is manifested. The classical studies with uniforms and robes evoke more radical shifts, imprinting a shift into strict categories (KKK-member, nurse, doctor). Similar studies have been carried out in virtual environments examining factors such as race or gender. Assignment of a dark skinned avatar has been shown to reduce implicit racial bias measured immediately after a virtual reality experience (Peck & al. 2013), and men will more frequently play female avatars when engaged in supporting gameplay roles perceived as feminine (Yee, Duchenaut & Nelson 2011 – although it's worth noting that the method of the study doesn't establish whether role choice is influenced by selection of a female avatar, or if avatar choice is preceded by the decision to play a supporting role; the study errs on the side of conservative, and posits the second).

Self-categorisation theory predicts that awareness of this categorisation according to avatar gender will alter the player's behaviour. This brings us to the first research

question: do players change their gendered behaviours in accordance to their avatars? If so, how do those changes manifest, and how can they be verified?

2.1.2 Examining assimilation through language

Virtual worlds come with two primary interaction mechanics: gameplay interaction (defined here as actions in the game interface proper) and language (defined here as textual interactions mediated through the game's chat interface, though some players also use voice chat protocols for the same purpose). This paper focuses on the second, as recorded in logs of the text passing through the player's chatbox.

To examine player chat as an indicator of assimilation, a coherent framework is required. Such a framework is suggested by Palomares & Lee (2009), in a laboratory study examining the linguistic assimilation of players to gender-matched or gender mismatched avatars. The dataset they examined was gathered through a specially designed trivia game, played co-operatively in CMC with a partner (supposedly human, in reality a simulated construct, whose responses were identical in each test). Human players were assigned randomly gendered avatars and asked to suggest correct answers, comment on their confidence levels regarding their answer, and finally comment on the success or failure of their choice. Their comments were then analysed for gendered language measured by a set of female-typical markers: references to emotion, apologies, and tentative language. The study found that male subjects assigned a gender mismatched avatar increased their usage of two out of three female-typical markers, while female subjects assigned a gender mismatched avatar decreased their usage of female-typical markers. The only variable that did not respond to avatar gendering was men's use of tentative language, which remained static regardless of the assigned avatar.

In their conclusions, the researchers note that the effect of avatar on language gendering in their sample is faint, and may not be replicable outside of a laboratory, or manifest similarly across all contexts. They also note their lack of metrics for measuring male-typical language: their method is restricted to testing assimilation to, and dissociation from, a female-perceived role, and cannot comment on whether a similar shift occurs in regards to male-typical linguistic behaviours. Both of these points are salient, if assimilation is to be examined in a natural environment.

Palomares & Lee carefully crafted their study environments to supply no gender cues other than those given by the avatar (by selecting a gender neutral trivia topic, fast food), and to gender avatars equally strongly to either gender (by having panels of responders rate their avatar options for maleness and femaleness). Commercial MMO's have none of these controls: rather, videogames as an activity are popularly considered a masculine environment, player avatars are designed to show strong gender diformity, with differing visual aesthetics for male and female avatars (see image 1). In an environment that strongly primes awareness of masculine identity and dimorphic presentations of gender, interactions are unlikely to be gender neutral.

This renders methods for evaluating male-typical language especially necessary to the experiment proposed in this study. If entry into a categorically masculine environment does obscure the more subtle gendering cues of avatars or player identity, arguably this would be apparent in relatively stable levels of masculine language, regardless of player or avatar gender.

2.2 Language and gender

For their set of female-typical metrics Palomares & Lee had a well-established set of operationalised metrics to employ. A corresponding set of male-typical features is conspicuously absent. To find a set of workable features, we turn to the scholarship on gendered language.

2.2.1 Female-typical language

The linguistic features Palomares & Lee use to categorise speech as female-typical hail back to the early work of Robin Lakoff, whose 1975 study *Language and Woman's Place* has been called the first systematic attempt to describe and develop methods for studying gendered language usage patterns. Two out of three metrics are directly from Lakoff's listing of features typical to women's language, and the third is implied by Lakoff's assertion that women are permitted more emphasis on emotional expression (Lakoff 2005).

Lakoff describes what she dubbed “women's language” in rather unflattering terms. Her list of features consists of the following: words for women's specific interests, 'empty' adjectives, rising intonation rather than declaratives, hedges, the intensive 'so', hypercorrect grammar, no sense of humour, and 'speaking in italics' (Lakoff

2005: 78-79). This mode of speaking is neither benevolent nor optional to the women schooled into it, she argues. The handicap built into women's language renders a woman less credible as a person – but refusing to use women's language renders her unfit as a woman, a social death-sentence (Lakoff 2004: 84-85). The double bind Lakoff describes is consistent with gender role theory, and fits the predictions of deindividuation and self-categorisation: adoption of the expected markers (women's language) to avoid social sanction (the stigma of unfemininity). It also shares similarities with Foucault's thoughts on normativity and normalisation, the process through which social structures become so accepted that they are considered not only everyday but the measure by which members of a group are judged (Foucault 2010).

Lakoff's study has been criticised on two significant counts: its introspective method (her theories were based largely on informal observation of her own circle of friends, and introspection on her observations) and her decidedly limited definition of femininity, described as gender essentialist and predominantly white, middle-class and heterosexual. The methodological criticism is justified, as is the criticism for depiction of femininity as a white middle-class monolith – but the accusations of seeing gender presentation as a product of some essential, authentic femininity seem at least partially misplaced. Rather, Lakoff argues that social structure precedes language, and that if the structure were different, the language would be as well (Lakoff 2004: 74). Women's language in Lakoff's writing is not an artefact of their sex, but a marker of their status as a marginalised group in relation to a dominant one. Similar forms of speech can be found not only in women's language but also with other groups removed from the power structures of hegemonic masculinity – groups such as male academics, counter-cultural hippies and gay men (Hall 2004:173-175).

Subsequent research has shown that the truth of gendered language is somewhat more complex than Lakoff posited. Quantitative studies have supported some of her ideas, such as women's higher use of hedges and interrogatives in the form of questions and tag questions as a form of conversation maintenance (Coates 1986:106, 117-118). Others, however, have been shown to be subject to manipulation by conditions beyond just the speaker's sex. Context can evoke apparently gendered speech patterns: linguistic analysis of witness testimonies in a North Carolina

courtroom found that women's language (defined as a set of features much like Lakoff's) correlates not with the speaker's sex, but with their social status and previous courtroom experience. Less status and less courtroom experience resulted in higher use of women's language, prompting the authors to forego the label 'women's language' in favour of 'powerless language' (O'Barr & Atkins 1980). In virtual chatrooms, gendered language seems to be tied to styles of communication: male-typical language is used to control and challenge, female-typical language to foster a sense of belonging and camaraderie, and both men and women will adopt strategies that convey courtesy or assertiveness as needed (Herring 2004). A study on weblogs found that the blog's topic, more than the writer's gender, controlled the bloggers' use of gendered language features: personal journals favoured a female-typical involved style, while informative aggregate-style blogs favoured masculine-typical informational language (Herring & Paolillo 2006). In both chatrooms and weblogs, it seems that gendered language correlates less with the writer's sex, and more with the social role they are performing.

As linguistic theory, *Language and Woman's Place* can be considered outmoded. It remains, however, a solid example of folklinguistic beliefs regarding gendered language, set out in terms of a linguistic model – and this, I argue, is far more relevant in pursuit of assimilation. For the cognitive processes of self-categorisation and assimilation, subjects will rely not on empirical linguistic data, but rather on their own folklinguistic stereotypes of gender-appropriate behaviour: and this has been demonstrated in at least some prior research. A popular folklinguistic stereotype is that women speak more than men do. A study on gendered language examined the verbosity of men and women in a standardised speech context (a Turing test identity game where anonymous men and women attempted to convince a panel of judges that they were a given sex). Men in the study produced consistently longer utterances than women did – and further lengthened their statements when pretending to be female – while women produced shorter utterances, and shortened them even more when pretending to be male. Even though women's actual statements were shorter, both men and women still performed femininity with more and masculinity with less show of verbosity (Herring & Martinson 2004).

The implication is clear: a firmly held folklinguistic belief on how a certain group behaves will cause speakers to emulate the fiction, not the reality, in their representation of the group. For purposes of this study, we can then suggest that if players assimilate to the gender of their avatars, they will emulate stereotypically expected patterns of gendered language, regardless of whether those patterns conform to reality.

2.2.2 Male-typical language

Studies on women's language have been criticised for their deficit approach – defining women's language in terms of what is missing. This makes it quite ironic that there are no tested metrics for quantifying the traits supposedly present in male language. There are a number of possible explanations for this. One candidate is the field of sociolinguistics coming of age alongside the shift away from second-wave, gender essentialist feminism to third wave, gender performative feminism: as perceptions of gender shifted from monolithic to socially constructed, it may have seemed a waste of time to go mapping the phantasm of 'men's speech' as a categorical construct. Alternately, Lakoff framing her model as a linguistic reflection of disempowerment may have pre-empted such investigation; in this paradigm dominant speech variants should be unconstrained, or at least constrained only by avoiding variants marked as marginalising. In this light it's understandable that studies on men's language have tended to a more particular focus – mixed-gender discourse, male-to-male discourse, or language indexing specific male identities, e.g. gay men or high school jocks. Finally, the scholarship seems increasingly to suggest that linguistic choice is driven by speech context and audience design. In such a paradigm, we're led to believe that recognisable gendering may be rendered only in those contexts where gender, specifically, is germane.

Still, some generalities can be drawn to suggest metrics for measuring stereotypically masculine uses of language. Some can be inferred from Lakoff's listing of features absent in women's language: assertiveness, dominance and the outright statement that women do not tell jokes (Lakoff 2004: 80) – but this list is poorly operationalised and vague. A more workable version is suggested in data-driven studies by Coates (1986). Her examination of women's language bears out many of Lakoff's findings, particularly for tentative language, references to emotion and some

topical preferences, and her examinations of male-to-male and mixed-gender discourse provide a plentiful list of features typical to men's speech.

In selection of operationalisable features of male-typical speech, two primary criteria of elimination were used. Speakers have been shown very adept at manipulating macrolinguistic features, e.g. gendered topics and vocabulary (Herring & Martinson 2004). They are however, less adept at manipulating more molecular features, e.g. apology and reference to emotion (Herring 2004). Useful metrics, then, should not be the most obvious topical and lexical categories, but more emotive features of stereotypical gender presentation. Likely candidates were selected based on how potentially fruitful they seemed to the case environment as a community of practice.

“Women don't tell jokes,” Lakoff claims (2004: 80): she leaves joking strictly in the realm of the masculine. Coates supports this, describing male-to-male interactions as thick with humour, particularly in the form of direct jokes, funny incidents and crazy stories. The ability to 'have a laugh', she says, is particularly important to young men, and “central to being acceptable as masculine” (Coates 2003: 53). This suggests direct humour as a male-typical metric, particularly given that the average MMO-gamer is a male in his early-to-mid 20's (ESA 2014:3), and that humour is one of the features used to index gamer identity (Shaw 2011).

Another candidate for male-typical speech is the oft-cited truism that men use more taboo and vulgar language than women do. Lakoff alludes to this, suggesting that women's language uses superpolite forms and hypercorrect content to couch indelicacy in euphemism and apology (Lakoff 2004: 80). Coates calls the quantitative claim mostly folklinguistic, but with some data to support it: a study found that men swear slightly more in mixed company, that women in single-sex groups swear more than men in mixed sex groups, and men in single sex groups swear three times as much as women in single sex groups (Gomm 1981, as cited by Coates 1986: 109). She argues that taboo language serves several functions in male discourse: it provides verisimilitude, conveys emphasis, and performs hegemonic masculinity by signalling toughness and manhood.

McEnery's (2005) examination of the British National Corpus provides a more detailed analysis of profanity in the UK vernacular. He divides taboo language into six categories: swear words (*fuck, piss, shit*), animal terms of abuse (*pig, cow, bitch*), sexist terms of abuse (*bitch, whore, slut*), intellect-based terms of abuse (*idiot, prat, imbecile*), racist terms of abuse (*paki, nigger, chink*) and homophobic terms of abuse (*queer, fag*). In McEnery's analysis, as with Coates, bad language is equally common with men and women, despite folklinguistic belief to the contrary. However, word preference is gendered, with a subset of words preferred by men (*fucking, fuck, jesus, cunt* and *fucker*) and another preferred by women (*god, bloody, pig, hell, bugger, bitch, pissed, arsed, shit* and *pissy*) (2005:29). He also describes a difference in the perceived strength of the swears preferred by men and women, with male-preferred swears considered stronger (2005:31) – and notes that while strong swears do index masculinity, both genders will contextually use mild nonswears (*fudge, dang*) to fulfil the same function as the swear words they replace. Finally, he concludes that swearing is used as a method of invoking distinction, to denote differences in both age, sex, social class and power (2005:50).

While swearing *per se* may be universal, the folklinguistic reading of it as masculine is strong, and the gendered divide shown by McEnery between softer and stronger swears suggests that there is a quantifiable difference in how swearing relates to the construction of gender. As with feminine-typical language, I argue that speakers assimilating a gender identity will be more likely to conform to usages they believe to be common among members of the group. As such, profanity as a metric for male-typical language seems germane.

Finally, according to Coates, much of male-typical discourse is shaped by the need to reaffirm hegemonic masculinity. This imprints on language in predictable ways, particularly the need to present as tough, unemotional and competitive (Coates 2003: 37) – a notable difference to Lakoff's women's speech, which is marked by the attempt to reduce conflict by the use of various hedges (Lakoff 2004: 79). Coates describes male competitiveness manifesting as the tendency to frame conversation in terms of heroism, conflict and achievement, to the point where storytelling itself becomes a competitive exercise, with 'speakers competing to boast of their triumphs or cock-ups' (2003: 37, 56). This introduces a stark difference to the types of stories

men and women tell in single-sex company: while women's stories describe communities of both sexes, men's stories describe lone male heroes facing down their enemies (2003: 137). The pattern persists to a lesser degree in mixed sex company. Though men will sometimes collaborate, or decline to challenge a dominant male to avoid censure, the performance of masculinity and avoidance of vulnerability is still central – and while women will sometimes subvert male display, they will do so more in the spirit of fun and collaboration than in competition (2003: 171-172). Herring observes a similar pattern in online mixed sex environments, noting that in CMC women's speech will more often foster camaraderie and solidarity, while men's speech will more often be sarcastic, boastful, or outright flaming (Herring 2004).

2.3 Corpus research into language and gender

All of the aforementioned research relies on an eyes-on examination of the material. Given that the data is generated in a convenient, digital format, it seems natural to consider options for automatic analysis as well. Human coding of data conveys some clear advantages, the most important of which is access to semantic and pragmatic data at lexical, sentence and discourse level. This comes at the cost of being fairly time intensive, as well as prone to bias where tokens are coded by readers privy to the conditions being examined. Corpus analysis methods, however, convey other benefits, some of which also offer potentially interesting avenues of investigation into gendered language.

Argamon, Koppel & al (2003) analysed single, double and triple n-grams of 604 academic and informal written texts from the British National Corpus, and used their findings to identify various syntactic markers that correlate with the author's sex. Rather than semantic details, their analysis focused on parts of speech and syntactic patterns found to correspond with authors of a given sex. In their findings, female writing was most distinguished by its use of pronouns. Female authors are likelier to use first and second person singular pronouns, as well as third person female pronouns; male authors use more plurals, third person male and the pronoun *its*. The same distribution is found for possessive, reflexive and basic underlying forms of each pronoun. The authors argue this corresponds to the idea of women's language as involved, described by Biber (1995): women's writing uses personifying pronouns to

encode relationships between author and reader or between actors in the text, while male writing prefers pronouns that depersonalise and diffuse individuality. Argamon & al (2003) also report that male writing has a greater tendency toward specifics, seen e.g. in heightened use of determiners, cardinal numbers and post-head modification of nouns with of-phrases, and a preference for adjective attributes. This they argue is in line with Biber's framework describing male writing as informational, concerned more with the particular details of the thing being talked about than its relationship to other things. The authors of the study claim an algorithm based on their findings can guess the gender of the writer of any given text with 80% accuracy.

While different in method, the underlying theory is similar to previous study of gendered language in one fundamental way: it takes as a given that gender is restricted to the male/female binary, and this bias is reflected in interpreting the data. Bamman, Eistensten & Schnoebelen (2014) question this assumption, as it risks smuggling predictive expectations of gender/language intrinsicness into results. In their analysis of a corpus drawn from 14,000 Twitter accounts, they perform two analyses. The first analysed PoS-tagged frequency data to establish the data's comparability with traditional binary gendered studies: of examined features, they find that pronouns, emotion terms, and CMC words are significantly linked to female gender, while numbers and technology words are significantly male-linked. Of words with mixed distribution in previous literature, they find emoticons and hesitations are female-linked, and swear words are male-linked.

The more interesting analysis, though, is the second one – a machine-learning based experiment mapping speakers into clusters based on similarities of word-usage. 14 out of 17 clusters showed a skewed gender split of 60/40 or more. Each group also correlated with several other intersectional indicators, such as race, age and affect, and was characterised by distinctive lexical choices suggesting a demographic (e.g. older women, businessmen, sports fans, African-American English). Examining the clusters for gender homogeneity, the authors found that the more homogeneous a network was, the more gendered the language used in the cluster was: the usage of same-gender markers increases with the homogeneity of the group. Similarly, speakers with gender-defying patterns were likely to have relatively few same-gender

connections. The authors argue that this shows how, in Butler's terms, gender is not simply a descriptive, but a powerful normative statement drawing lines around who is included and who is excluded. While the first analysis shows the study data performs in line with accepted literature on how language is gendered, the second analysis shows that it does so in far less homogeneous terms than analysis as a strict binary may suggest – and that with the flexibility allowed by computational methods, richer treatments of gender may be in order.

While novel in a quantitative study, this sentiment is not new in the study of language and gender at large. From an approach that once emphasised differences, attention has increasingly shifted to one that emphasises similarities, and the ways in which language is no longer the natural product of essential gender, but rather a creative strategy for indexing gender as one of many aspects of social identity.

3 Methods

This section describes the research questions, the case environment, the data collection procedure and the operationalised metrics used for coding the data. Also addressed are research ethics and choices regarding statistical analysis.

3.1 Research questions

The questions this study attempts to answer are these:

- 1) What kind of differences can be found in players' use of gendered markers?
- 2) Are these differences tied to the player's gender, or the gender of their avatar?
- 3) What other factors may explain observations?
- 4) How could the study design be refined to better answer these questions?

3.2 The case environment

World of Warcraft, commonly abbreviated as *WoW*, is a fantasy videogame of the Massively Multiplayer Online Roleplaying Game genre (abbreviated as MMORPG or MMO), played on officially hosted servers that players pay a monthly subscription fee to access. Each player is represented by an avatar with a set of customizable

features: faction, class, race, sex and appearance. The character's class (e.g. warrior, mage, hunter) determines what game mechanics and group roles are available to them. Choice of faction (Alliance or Horde) determines which races are available to a player, and choice of race (e.g. human, elf, orc, troll) adds minor gameplay perks. The sex and appearance of a character are considered cosmetic and have no impact on gameplay performance – though the cognitive theories of assimilation described earlier suggest that they do have an effect on players' social performance.



Image 1: Concept art of the male and female versions of various player races in World of Warcraft, demonstrating the heavy dimorphy of gender portrayal in the game. Male avatars show clearly more variety of body shape and posture than their female counterparts. The difference is most obvious with the Troll models (bottom), with the male shown as betusked, slouched and lanky: the female, meanwhile, is much the same slim and conventionally attractive shape as the Human and Elf models. It's also worth noting that the most conventionally beautiful male avatar, the male Blood Elf model (middle), was redesigned to be more muscular before release, due to player complaints of it looking 'too gay'.

The core gameplay of *WoW* centers around levelling up the character and group activities called dungeons (for groups of 5) and raids (for groups of 10-25 players). For easier access to the group content, most players will join a guild – a permanent, player-run peer group of people who regularly play together. Guilds will vary in terms of size (from single digits to hundreds), intimacy (from close friends to strangers), expected time commitment (from casual logins when time permits to several hours of focused play daily) and their activities of interest (e.g. levelling and questing, player against player combat, dungeon or raid content, or roleplay between players – the last three items are abbreviated in gamer parlance to PvP, PvE and RP, and will in this study be referred to collectively as tactical content).



Image 2: The game interface of a player engaged in casual questing gameplay. The chatbox, where the data of this study is collected, is at the bottom left of the screen.

In such a social environment, communication with fellow players is key to success in the game. The basic mode of communication is an ingame text-based chat box implemented into the game's graphical interface, similar in functionality to instant messaging applications such as MSN, Skype or Facebook Messenger. Messages in the chatbox are restricted to 250 symbols in length and the ASCII-symbol set; by default chat is set to spatial (the avatar's immediate surroundings in the gameworld), but the player can also choose to speak to their guild, party, instance, raid or various

other channels, including custom channels shared by friends and whispers/tells addressed directly to a single player.

By default, the player's game client records all text going through their chatbox (or chatboxes, as the interface can also be modified to show several separately threaded conversations at once). This is the data the study focuses on. All interactions are also recorded serverside by the game provider: but this data is not generally available.

3.3 Research ethics

Analysing logs of everyday interactions via ingame CMC presented some ethical considerations. They were addressed as follows:

Informed consent. Participants were told their data will be used for a study dealing with virtual environments and the things that shape the language gamers use. Examined variables include e.g. avatars, gender, age, education, favoured gameplay activities and gameplay hours. All responders were promised anonymity, and given the option to gender themselves freely (open text rather than a forced binary male/female choice) – all self-reported as male or female.

Privacy and safety. Logs were submitted under the player's avatar name. All private messages (defined as messages not sent in an open channel) were discarded prior to coding, as was any personally identifying information (names, e-mails, RealID, social media handles, etc.) remaining elsewhere in the logs. Original avatar names were recorded for cross-referencing a player's gameplay activity with *WoW*'s public Armory database; however, in the samples cited in this paper, all names of players and avatars have been replaced with randomly generated placeholders.

Minors. While all responders reported their age as over 18, the possibility remains that in a T-rated game (suitable for age 12 and above), minor responders may surface in the larger sample. As all chat is generated in the course of normal gameplay activity, and concerns of privacy and personal safety are observed for all responders, participation in the study does not pose a risk to minors.

It is acknowledged that obtaining informed consent risks contaminating the data in terms of alerting participants to gender as a salient variable, always a risk in cognitive research. This was deemed a more ethical alternative than subjecting

responders to normative gender evaluation without their consent. The risk of contamination is hopefully minimised by listing gender as one of several tracked variables. Also, as chatlogs are generated with time coding and alongside normal gameplay, in many cases it was clear that the logs submitted had been recorded long before the data request, thus eliminating the possibility of contamination altogether. Where possible, such logs were preferred over more recent ones.

3.4 Data collection and sampling

Data for the pilot study was collected from a single guild on the European server Argent Dawn, to minimize variation due to differing communities of practice.

Volunteers were asked to submit the following, via open text field:

1. chatlogs collected in the course of at least one evening's normal gameplay activities, playing one or more avatars
2. the name and server of the character(s) chat had been logged for
3. a list of their preferred gameplay activities (RP/PvE/PvP/social/etc.)
4. estimated weekly gameplay hours
5. demographic data (age, gender, nationality, first language and education)

Using the character/server names, additional information on the player's avatars and gameplay practices was retrieved from *World of Warcraft's* Armory website. The Armory automatically tracks all player avatars, listing features like an avatar's race, sex, power, community memberships and significant gameplay milestones such as items acquired, enemies killed or challenges completed. For this study, Armory data was used for gendering player avatars, verifying guild standings, and as a control when rating players' self-reported interest in tactical gameplay as competitive (avatars with rated PvP listings and/or progression in current raid content) or casual (avatars with neither).

Before analysis, logs were scrubbed of text from sources other than the player who had logged it (e.g. system messages, other players), as well as all private messages (RealID, BattleTag, whispers/tells). Logs were then compiled by avatar, and pre-examined for viability. Logs were deemed unviable for two reasons: roleplaying content and/or insufficient length per avatar. Some logs contained large quantities of roleplaying chat: this was deemed unsuitable for analysis (as described in section

4.1.1 Logged speech contexts). Exclusion of roleplay chat caused substantial data loss, as for some players it constituted as much as two thirds of the log. Logs of insufficient length were either the result of discarding roleplaying content, or logged by players engaged in casual PvP or PvE activities with strangers, where familiar gameplay rendered communicative needs minimal and the company of strangers discouraged casual chat. All discarded logs were at most 10 lines in length (approx. 50 words or less).

After exclusion of roleplay chat and very short logs, 6 out of 12 responders had provided usable logs. Two responders provided samples on both a matched and mismatched avatar. This made a total of 8 logs – two for each examined condition (male-matched, male-mismatched, female-matched, female-mismatched).

Group of responders

Player sex	<i>n</i>	Nationality	Activity	Playtime
Responders for all logs (<i>n</i> = 6)				
Male	3	Finnish, Swedish, Bulgarian	Social (3), Tactical (2)	20+
Female	3	Finnish, Finnish, Norwegian	Social (3), Tactical (2)	20+
Responders for paired logs (<i>n</i> = 2)				
Male	1	Bulgarian	Social	20+
Female	1	Norwegian	Social, Tactical	20+

Table 1: A summary of the responders and their demographic data. Mother tongue matched nationality for all responders. Self-reported player activities were combined into two labels, Tactical (for competitive activities reported as PvE, PvP, raiding and dungeons) and Social (for social activities reported as RP, chatting, social and casual). All responders reported 20 or more hours of weekly gameplay time, suggesting high investment in the game. Education (50% university students, 50% other) and age (26-35) were also requested to mask the salience of player sex as a variable, but not tracked in analysis.

Logs included for analysis varied in length from 311 to 2203 words, or 67 to 325 lines. For standardised analysis, two standard samples were cut from the beginning of each log: 67 lines (according to the shortest line length, for analysing verbosity), and 311 words (according to the shortest token count, for most analysis). Lines here are a measure of length native to the *WoW* chat system, a single message no longer than 250 symbols. Excessive numbers of lines per minutes are blocked, to prevent

chat abuse in the form of spamming: this makes line length a more reliable measure of verbosity than frequency of messages (though this could also be tracked in the logs, via time codes).

The total wordcount of final samples analysed is 3676 words for the 67-line verbosity sample, and 2488 words for the 311-word sample used for other analysis. This is an extremely small sample, both in terms of wordcount and number of responders. Though small, the dataset is balanced for gender. Responders ($n = 6$) self-reported as 3 male, 3 female; paired samples ($n = 2$) were logged by both a male and a female player. Two players reported exclusively social gameplay interests, while 4 also had an interest in competitive PvP/PvE activities. All responders were fluent second language speakers, listing their nationalities as Finnish, Swedish, Norwegian or Bulgarian; there were no native English speakers in the sample.

3.5 Operationalised metrics and coding

After the logs had been anonymised and standardised, they were coded for operationalised metrics. Where a phrase could fit two categories, it was filed into both. While this produced a somewhat higher incidence of tokens, a symmetrical increase across all data was deemed preferable to possible selection bias by forcing a choice between categories.

Previous studies on gendered language in CMC have focused on female-typical metrics sensitive to gendering. For comparability, the pilot study adopted the Lakoffian metrics used by Palomares & Lee: references to emotion, apology and tentative language.

To complement the female-typical metrics, a list of male-typical metrics was compiled. The operationalised set of male-typical features selected for this study was based on features suggested by Coates (2003): taboo language, competitive expression, and direct humour. These were further refined by reference to McEnery (2005), documenting profanity as male-typical though not male-exclusive speech. Selection was informed by folklinguistic belief in a feature as gendered, and by clear ways to operationalise the metric.

The metrics were operationalised and coded as follows:

- I. **Reference to emotion** is defined as any mention of an emotion (e.g., I love

my selfheal, that's terrifying, I'm so confused).

- II. **Apologies** are defined as a direct statement of being sorry (e.g., sorry, my bad).
- III. **Tentative language** is defined as the combination of language features that indicate uncertainty and low confidence: *hedges* (e.g., I think, kinda, maybe, probably), *disclaimers* (e.g., far as I know, I'm not sure, don't quote me) and *tag questions* (e.g., don't you think? isn't it? Right?).
- IV. **Taboo language** is defined as swear words (e.g. shitty, dammit, ffs etc.), insults (you're a thick one, bitch) or vulgar/taboo topics (sexuality, bodily functions, sexism, racism, bigotry). Softened nonswears (faffed it up, crap) were excluded.
- V. **Competitive language** is defined as reference to heroism, competition and struggle (e.g. he's got serious skills, challenge yourself, up for something difficult?) or to victory/defeat (e.g. it's ours now!, she one-shotted me).
- VI. **Direct humour** was defined as puns, jokes, funny anecdotes and reference to shared cultural capital like repurposing game features (/copycat), internet memes (such wow) or pop-culture punchlines and quotes (I'm your father, Luke).

Though the coding was primarily sufficient, a few isolated ambiguous cases were found. These were left out of analysis for this sample, but recorded in the discussion section for review in use with a new sample. While funded studies have the luxury of using research assistants or services like Amazon Mechanical Turk to facilitate this, for this study, no outside coders were used.

3.6 Corpus based analysis

Manual coding has two primary faults as a research method, both stemming from its reliance on human coders. First, as sample sizes grow, it quickly becomes slow and labour intensive. Second, it is prone to bias based on the perceptions of the individual

coder: commonly this is controlled for by using at least two independent coders, ideally blind to the condition being examined for.

As an attempt to escape this limitation, some corpus based methods of analysis were also explored. The standardized samples were POS-tagged using the Carnegie-Mellon ARK Part of Speech tagger (Gimpel & al 2011). The tagger is intended for Twitter English, but *WoW* chat is fairly similar as a speech context: informal CMC messages of limited length between semianonymous users. Some trials with POS-taggers optimised for written language (Ubiquity and Stanford POS-tagger) bore out this assumption: where other taggers produced a high ration of unrecognised or mistagged tokens, ARK tagger performed with a high level of reliability, erring mostly where common nouns were used as proper names.

The gender-typical features examined in the manual coding rely heavily on a semantic component, and as such are difficult to access through corpus analysis. However, corpus methods have been used to suggest other gendering features. A subset of features reported as gender-linked and accessible by POS-tagging in an online environment was selected, based largely on Bamman et al (2014) for ease of comparison. These included

- emoticons and pronouns (reported as female-linked), and
- numerals (reported as male-linked).

3.7 Statistical testing

Basic t-testing was done to test the correlation of measured features with player gender and avatar gender, and of correlation of a feature to male or female player gender with a gender matched or mismatched avatar. The test was chosen due to limited data, and acknowledged as problematic due to the application of a parametric test to data that may not conform to a normal distribution. On a sample large enough to allow replication of Palomares & Lee's method in full (ANOVA), this should be done.

While significance testing can theoretically be done on samples as small as $n = 2$, samples smaller than at least a few dozen datapoints are especially vulnerable to the influence of outliers; this can cause both the low and the high extremes of possible range to be overrepresented. Methodical prudence demands that extraordinary claims

deliver extraordinary evidence; results from a sample this limited should be treated as highly provisional, regardless of how well they fit the model.

4 Data and findings

Answering the question of whether assimilation occurs is beyond the scope of a pro gradu. Instead the pilot study set out to do two things: first, establish that the study is possible to conduct on the available data (chatlogs, their accompanying questionnaires, and avatar-related data mined from *WoW Armory*), and second, to test the method on a limited sample to discover possible weaknesses or oversights in it, so that the method might be refined accordingly before applying it to a more representative sample. Of special interest was evaluation of the male-typical metrics set: first to see whether it actually produced useful measurements, but also because research into gamer subculture suggests that gamer identity is indexed as predominantly male (Shaw 2010; Shaw 2011), a fact likely to reflect in the language used in a gaming context. Since different gameplay contexts also generate different social norms and pragmatic pressures on language (e.g. descriptive prose style when roleplaying, or short-hand brevity in hectic combat), it was deemed possible that gameplay contexts might be reflected in the collected data, and in the way gender was encoded in them. To allow a more reliable estimate of a player's investment in a particular gameplay activity, and by extension the type of player identity they might index, responders were asked to name their favoured gameplay activities, and their Armory profiles were checked for group memberships and/or recent achievements in competitive gameplay.

Given the small sample size, large variability in the data was both expected and found. The analysis attempts to disentangle variance due to sampling or context from variance attributable to the intended metric – however, all analysis drawn from such a small set of data is provisional at best, and best used mainly to further more rigorous experiment design.

4.1 Data

4.1.1 Logged speech contexts

Chatlogs record all the text passing through a player's chatbox (or chatboxes, as the interface can be customised to allow for several), including player-written messages,

system messages (e.g. log on /log off messages), game feedback (e.g. warnings on monster abilities, taxi services etc.) and immersive elements (dialogue from computer-controlled characters, environmental alerts). For the purposes of this study, only player-written messages were examined.

Player chat fell broadly into three speech contexts. The first context was coordinating gameplay with a group of peers, usually in or in preparation for one of the game's many tactical gameplay challenges. Tactical chat focused on the gameplay situation at hand, including organizing groups or discussing strategy, or notifying group mates of temporary pauses or interruptions to gameplay.

Taja: how's you May, any chance I could drag you along to get repeatedly stomped over in a challenge mode?

Taja: (we'd die, brutally, many times)

Taja: there's no Anna online at this moment, so no hurry :)

Taja: I'm thinking Anna, you, me, poke Ben to grab Brildur.. and then someone deepsing.

Taja: preferrably ranged, and preferrably with good CC abilities.

Rilma: Take the first door when we get upstairs and clear that whole room.

Rilma: Not.. not in one pull.

Rilma: <.<

Rilma: All set?

Rilma: Chaeerg

Rilma: Feel free to pull the next by the time the last mob dies~

Anendil: Im getting some coffee, 5 mins.

Anendil: Back

Zenji: brb juggling cats

The second context is best described as social: discussion unrelated to the game, such as personal disclosures, exchanges of opinion and discussion of popular books and films. Social discussion primarily took place in extended pauses in tactical gameplay, or in spatially unconstrained channels alongside solo gameplay. Though not directly related to the undertaking at hand, social chat was also frequently flavoured by the game environment, e.g. by reference to game mechanics: the slash-command format is used to make avatars perform scripted actions such as dancing or saluting; and building reputations with various factions is one of several forms of advancement in the game.

Zabu: workday was alright but i managed to get some sort of post-work rage on my way home

Zabu: i'm sure it'll get better soon

Galjin: i want bandages too

Galjin: /copycat

Crakar: Have you seen The Empire Strikes Back?

Crakar: The one where Luke kisses his sister.

Crakar: No, that's Bulrug. You've not been paying attention.

Crakar: And thus you lost 1000 reputation with me. :|

The third gameplay context, which in this sample formed an unexpectedly large part of collected data, was roleplaying – a creative exercise in which players make an effort to portray their characters as fully developed fictional personas. Most roleplay takes the form of a dramatic fiction narrated in public channels, alongside the player's nonfictional interactions as themselves in other channels:

Galjin looks slightly disoriented, the whining gets louder as he tries to move his arm. Eventually he manages to lean back to his original position, although it is a very slow process.

Galjin tries to reach for the waterskin, but ends up whining instead. "... wha' happened? did we win?" His voice is quite hushed.

Galjin says: ... I don' feel like a winnah

[Guild] Galjin: i'm listening to rihanna. i wonder should i go and see a doctor

Taja says: This is warning, blue-eye.

Taja takes the bottle with her left hand

Taja says: You touch wards, there is curse. You touch Bu'jachi, there is curse. Witch things, not for blue-eyes.

[Guild] Taja: oh, must've missed that while laughing at Dora. :P

Taja grins amiably at Isadora

Taja says: I warn first, eh? Next time, real curse.

As the excerpts show, roleplayed chat differs dramatically from the other two logged contexts. Galjin's first two lines are written in a formal narrative prose style, followed by a line of direct speech written in a simulated accent, and an unrelated joke to Guild chat, to players not participating in the roleplay scene. Taja's direct speech lines show a similar fake accent, the descriptive lines are written in standard English, and the metacomment in Guild chat reverts to casual internet English, complete with dropped capitalisation and an emoticon. This shows three different layers of interaction between players, each coded into shifts of register and chat channel. The messages to Guild are standard social chat: the messages in spatial speech (the fake accent lines) and spatial emotes (the standard English lines) are elements of roleplay.

Online roleplay is an improvisational storytelling exercise where players agree to use some channels for interacting “in character” (IC) – that is, pretending to be the Orc Warrior or Gnome Mage their avatar shows. In channels left out of the roleplay agreement (here represented by Guild chat), they still speak “out of character” (OOC), as gamers playing *World of Warcraft*. IC and OOC gameplay can occur in parallel in different channels, or a player might focus on only one type of interaction at time – in either case, due to its mimetic component and stylistic idiosyncracies,

roleplayed chat is easily identifiable in the logs. It is also worth noting that gender-swapping and roleplay are overlapping but separate phenomenon. Gender-swapping is creating an avatar of mismatched gender and ellipting the player's sex: roleplaying is creating a fictional personality and life-story to go with the avatar, and performing as that person to other players in channels designated as IC – in OOC channels, the player of the Troll armsman is still free chat away about her cats, knitting hoddy or a lousy day at work.

While speakers attempting to alter their gender presentation have been shown to target primarily topical and vocabulary cues, rather than more molecular ones like emotive or tentative language (Herring & Martinson 2004), the interest of this study is on subconscious influence, not active impersonation. Deliberate mimesis is a potential source of additional variation: for this reason alone, roleplayed content is less than ideal as a source.

Apart from the problem of mimesis, manual coding of the pilot sample also showed that roleplay impacted logged language in at least four ways. First, roleplayed chat differs in terms of textual features: acronyms, emoticons, shorthand and other CMC'isms were absent from roleplayed chat. Second, linguistic register shifted from casual internet English to something closer to formal literary prose. Third, the stylistic shift to a dramatic narrative style changed how emotion was expressed: roleplayed chat was much denser in emotional expression. Fourth, roleplay conventions of how actions are expressed promote a higher level of tentative expression. In popular *WoW* roleplaying style, actions against other players are described as open-ended, to allow targeted players the choice of going along with the action or countering it: this means that a provisional expression, “She starts casting the spell, and might hit anyone who stays close to the target,” would be preferred over a definite expression, “She casts the spell and hits anyone near the target”. Between the formal literary style and the increase in emotional and tentative tokens (both gendered as feminine markers in Lakoffian linguistics), roleplayed chat stands out from chat collected from non-roleplaying contexts. Logs containing primarily one or the other were rendered dramatically different.

Since roleplay chat is both theoretically problematic (deliberate mimesis) and qualitatively distinctive (promotes certain gendered modes of expression over

others), it was ruled out of evidence. Analysis was restricted to the out-of-character gameplay contexts (tactical and social context), where players interact as players of *World of Warcraft*, rather than as imaginary inhabitants of Azeroth.

Speech context shaped both the quality and quantity of chat logged. Social chat was logged primarily in the Guild channel, a closed channel shared between players and the peer group they play the game with, or in the Party and Instance channels, which allow communication between the players currently grouped together. After the removal of roleplay chat, this constituted the majority of logged text.

Tactical logs were, for the most part, brief. The trend in *WoW* since its inception has been to lower the threshold of participation in group activities. Where in days past players would assemble groups of friends or strangers one by one and plan dungeon strategies as they went, currently (with *Warlords of Draenor* as the most recent expansion) group assembly is mostly handled automatically, via the Dungeon Finder and Raid Finder autogrouping features. Likewise, most gameplay content has a default solution that players can easily execute without prompting or coordination: more communication is required only if the group lacks sufficient power to attempt the default strategy, or if the group is attempting a more difficult strategy (which generally also yields more rewards).

Most tactical content in the sample seems to have been run on default strategies, as the logs produced from them are very brief. Only three logs included use of the Instance channel, the default channel for tactical undertakings. One recorded an experienced player instructing their team in an uncommon, more difficult strategy, analysing failures and correcting tactics for better performance:

[Instance] Rilma: She aggroed me

[Instance] Rilma: for like a split second

[Instance] Rilma: Oneshot me.

[Instance] Rilma: BL set for this'un

The other two were typical to a stock tactics run – half a dozen lines consisting of a greeting, minimal smalltalk, and CMC shorthand or requests to deviate from the stock tactics.

[Instance] Elvania: Heya

[Instance] Elvania: ^^

[Instance] Elvania: Have you seen our gear?

[Instance] Elvania: You are a thick one,aren't you?

[Instance] Elvania: Thanks for the run!

[Instance] Elvania: Could we kill the spider as well? I wouldn't mind the extra chance for some armor tokens

While chat logging remains viable as a data gathering method, it now seems to increasingly target established social groups. Most fruitful channels for analysis seemed to be Guild, Instance and Party chat channels, though isolated instances of Raid, Global, Trade and LFG were also included. Spatial chat in this sample was primarily used for roleplay, though this was likely a feature of the server community: or a server without a roleplay designation, spatial chat would likely be equally usable.

4.1.2 Sample length

To examine the impact of sample size, all logs were analysed twice. Once in full, to determine optimal sample size for catching variance in the data, and a second time as standardised segments cut from the beginning of the log to match the length of the shortest complete log (67 lines for the verbosity analysis, 311 words for all other analyses). Other studies have found that the threshold for accurately estimating gendered language use lies around 400 words (Argamon & al 2003). Unfortunately, the elimination of roleplayed data dramatically shortened some of the logs, and so this study had to make do with less. Unabridged logs ranged in length from 311-2203 words, or 67 and 325 lines. Line lengths are included largely for the purpose of measuring verbosity, and because they form the native standard of message length in *WoW*. Standardised samples were deemed prudent for closer analysis, because the length of a log was not tied to the length of its lines: the log with the least lines (67 lines, 478 words) was over a quarter longer than the next shortest (69 lines, 311 words).

Close examination of unsampled logs provided evidence of the obvious: that a player's voice will respond to context, sometimes dramatically. A male player on a

matched avatar, who in the truncated logs coded as very clearly masculine (low on all the metrics, but higher on the male-typical ones), produced two 10-line clusters in his unabridged log where almost every line contained tentative language in the form of hedging. The segments come into focus through context, and through examination of the avatar's Armory data, which identify the avatar as the leader of their guild. In both cases, the spike in tentative expression is tied to the player welcoming a new recruit to the guild and advising them on guild policy. The brief adoption of female-typical hedged expression instead of the player's usual high levels of male-typical competitive language echoes Herring's (2004) observation that in CMC environments, both men and women will contextually adopt strategies that convey courtesy or assertiveness, regardless of how culture genders those strategies. Male-typical language will be used to control and challenge, while female-typical language will be used to foster a sense of belonging and camaraderie. A guild leader's deliberate shift to a feminine register when greeting new recruits is well in line with this expectation.

The example shows that videogame chat is no less sensitive than its realworld counterpart to sample size. Brief segments are prone to colouring by momentary assumption of specific roles – whether that role is a fully formed fictional character, or a linguistic community building device. For future projects, this should be observed, and data collected with a firmer eye on the viable minimum sample.

4.2 Findings

Palomares & Lee (2009) suggested that women seem more sensitive to manipulation through the visual cue of avatar appearance, but men also respond to visual priming by increasing their use of tentative language when given a female avatar. This study partially supported their findings. For female-typical metrics, men did use more emotional and tentative markers on a female avatar, but the increase was statistically significant only for tentative language. For male-typical metrics, both men and women used more humour while playing a male avatar, but the difference was not statistically significant. Taboo language (male-typical) and apology (female-typical) produced only very low token counts – a range of 0-2 per log – and as such could not be meaningfully analysed.

A correlation between avatar gender and verbosity was also observed, in terms of players assimilating to the folklinguistic belief that women speak more than men –

both men and women logged more average words per line when playing a female avatar. While interesting, the shift remains narrowly below the traditional threshold for statistical significance.

Emotional language and competitive language defied expectations of gender for both player and avatar-based analysis, particularly for female players – female players used less of both emotional and tentative expressions when playing a female avatar. This is most probably explained by a combination of the speech context and the players' interest in high-end competitive gameplay (as determined by self-reporting, and completion of prestigious gameplay content recorded in the avatar's Armory profile) – a significant correlation was found between interest in tactical content and use of competitive language. This shows a need to adjust the experiment design to allow for gameplay context and/or the player's areas of interest in the sampling. This is discussed in more detail in section 5.1.1.

Summary of correlations and significance testing

Tracked metric	Observed correlation	Significance
<u>Gender (n = 8 logs)</u>		
Verbosity	MM/MF - increased with female avatar	ns
Emotion	Female players	0.0128
	FF/FM - increased with male avatar	0.0198
Tentative	Female players	ns
	MM/MF - increased with female avatar	ns
	Paired samples - increased with female avatar	ns
Humour	MM/MF - decreased with female avatar	ns
	FF/FM - increased with male avatar	ns
<u>Social / tactical activity (n = 8 logs)</u>		
Competitive	Female avatar	0.0129
	Tactical gameplay context	0.0333
Emoticons	Tactical gameplay context	0.0179

FF/FM = female player, female avatar / female player, male avatar
MM/MF = male player, male avatar / male player, female avatar

Table 2: A summary of the observed correlations and significance of the correlation. All data was analysed for correlations to player gender, avatar gender, and the matched/mismatched avatar condition for male and female players. Also listed are all correlations that conformed to the model in raw data, but not at a statistically significant way.

Of particular interest were two paired samples, where the same player had logged chat from a male and female avatar. While no statistical conclusions can be drawn from two datapoints, paired samples control for many of the variables otherwise contributing noise (e.g. player interests, attitudes and idiosyncracies). For future analysis, this is the preferable set-up for sampling.

4.2.1 Verbosity

Statement length has been shown to be sensitive to gender (Herring & Martinson 2004). This was the first metric tested, by means of average line length. After discarding roleplayed chat, the shortest logs measured in at 67 lines (478 words), or in words 311 words (69 lines). Average line length was determined by an automated wordcount of the first 67 lines of all logs, divided by the number of lines.

The line-length sample was taken to examine the hypothesis based on verbosity. In the line-count standardised logs, male players produced a total of 2085 words, female players 1591 words. However, male avatars produced 1680 words while female avatars produced 1996 words. This is consistent with previous findings – though men de facto use more words in CMC, the popular prejudice that women are chattier will make men increase and women decrease verbosity on a gender swapped avatar (Herring & Martinson 2004).

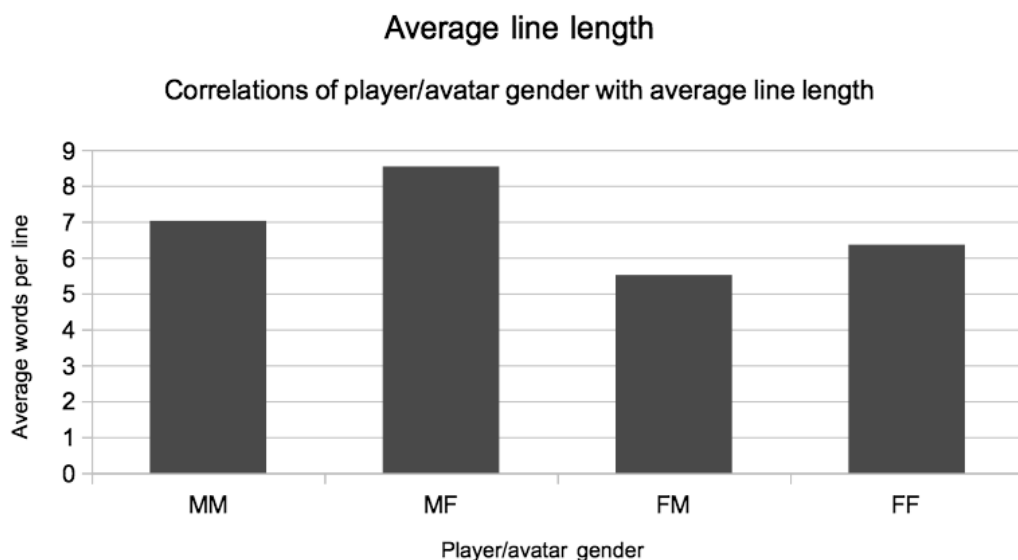


Table 3: Verbosity, measured as average words per line, in the examined combinations of player/avatar gender.

As predicted by previous research on male and female verbosity, women in the matched condition (FF) use less words than men in the matched condition (MM). As predicted by the assimilation hypothesis, the mismatched conditions of each gender replicate the folklinguistic belief appropriate to the avatar: that women speak more (MF) and men speak less (FM). Though in line with the assimilation hypothesis, the shift was not statistically significant.

Previous findings were supported in the raw data, but not in statistical testing. No significant correlations were found between line length and player gender or avatar gender. The nearest result to significance was male player verbosity increasing when swapping to a female avatar, but the result ($p = 0.0771$) did not pass traditional thresholds for statistical significance ($p < 0.05$).

4.2.2 Female-typical language markers

Manual coding examined metrics based on three female-typical features used by Palomares & Lee (2010), as well as three male-typical features derived from previous research (Coates 1986; Coates 2003; McEnery 2005) to complement them. Female-typical markers consisted of reference to emotion, tentative language and apologies.

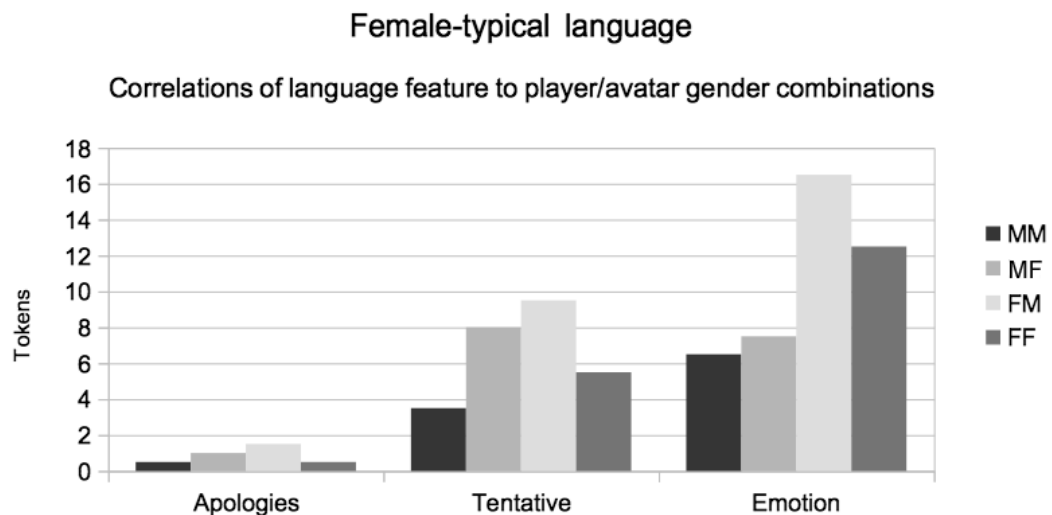


Table 4: Female-typical language across observed combinations of player/avatar gender, shown as average token count per condition. Male players in both matched (MM) and mismatched (MF) conditions show the expected assimilation pattern. Female players support previous research by scoring higher than men in matched condition (FF) but contradict the assimilation hypothesis by scoring even higher in mismatched condition (FM).

Apologies were logged at very low frequency (0-2 tokens per log), and showed no significant correlation with player or avatar gender, or as a matched/mismatched avatar pair for the same players. Given the shortness of the samples and the extremely low token counts for the metric overall, it is likely that tokens were simply insufficient to capture variance. It may also be that in gamer parlance direct apologies are marked language use, and as such generally avoided. In the examined logs, mistakes were usually acknowledged by the mistakee by interjections (*oops*, *derp*) or admission of fault without actual apology (*I didn't see that*). Direct apologies or statements of regret (*sorry*, *my bad*, *apologies*) were few and far between.

Tentative language was found most in line with expectations. No correlation was found between tentative language and players' reported gender. Male players in the matched and mismatched conditions (MM/MF avatar pairing) showed some correlation between female avatars and increased use of tentative language; however, the correlation remained below statistical significance ($p = 0.087$). Controlled comparison of two players (one male, one female) playing both a matched and mismatched avatar showed a correlation between female avatars and increased tentative language ($p = 0.057$): this skirts the threshold of significance, but doesn't quite pass it. It is mentioned to track a potential point of interest pending a more representative sample – but as Palomares & Lee note, the faintness of assimilative evidence in even lab conditions is faint, suggesting that in a natural environment, the effect may well be fainter still.

Emotional language proved the most surprising metric. Only lexical tokens were included as emotional language by Palomares & Lee (2009), most likely because their test environment specifically prompted formal language statements. *WoW* chat, by contrast, routinely employs various CMC-specific devices, and of those at least emoticons (Wolf 2000, Bamman 2014) and interjections (Acton 2011, Bamman 2014) have been reported as shorthand for emotional expression in CMC-based environments. Coding for emotional language was done on three levels: lexical (*rage*, *happy*, *confused*), interjections (*yay*, *awww*, *haha*) and emoticons (^, :D, >.<). Token counts for these three measures were combined into a composite measure for emotional reference. The composite measure showed a significant correlation between emotional language and female player gender ($p = 0.0128$ – the

only case of a metric significantly correlating with player gender in this sample), as well as a significant and surprising decrease of emotional language on a matched avatar ($p = 0.0198$). In the sample dataset, female players on male avatars (FM) used more emotional expression than female players on matched avatars (FF) – not less, as the assimilation hypothesis predicts. This finding is further examined in 4.2.4 Gameplay context as a shaper of language.

4.2.3 Male-typical language markers

Male-typical markers, as to be expected with an experimental method, were a more mixed bag. The set proposed in this study consisted of taboo language, competitive language, and direct humour.

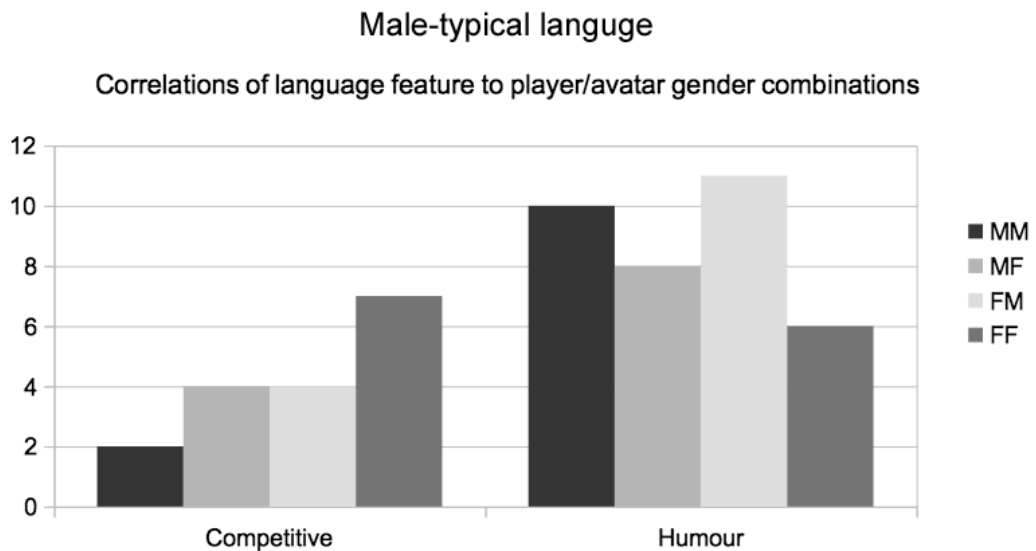


Table 5: Male-typical language across observed combinations of player/avatar gender, shown as average token count per condition. Humour performs as predicted by the assimilation hypothesis for both male and female players, with both male and female players using more humour on male avatars (MM and FM condition, respectively). Competitive language defies the assimilation hypothesis, showing both male and female players use more competitive language on a female avatar (MF and FF condition, respectively).

The token count for **taboo language** fell below meaningful analysis levels. A single case of *shit* was logged by a female player on a female avatar, and two cases of softened non-swears (*faffed it up*, *crap*) by male players on a female and male avatar respectively, as well as a single case of insult to intellect by a male player on a female avatar (*You're thick one aren't you?*). This is contrary to expectations – previous research would lead us to expect nonswears from women and their harder

counterparts (McEnery 2005), and to expect gamer parlance in general to be predisposed to masculine coding (Shaw 2010). A likely explanation is due to all responders belonging to the same guild. MMO players create subcommunities with shared standards on what they consider desirable language in that particular community, and peer-based monitoring to ensure the policies are followed (Collister 2014). Since the bulk of the data here comes from the community's Guild-channel, community language standards are likely to have kept profanity levels lower than they would be in unmoderated, cross-community contexts.

Direct humour performed entirely as predicted by the model in raw data, with men logging more humorous chat than women on matched avatars (MM, FF), and men decreasing their usage on mismatched avatars (MF) while women increased theirs on mismatched avatars (FM). However, none of the shifts were statistically significant; nor was there a systematic correlation with avatar gender or players' reported gender. In this, humour echoes the performance of the female-typical metric for tentative language: while the expected pattern is there in this data set, it needs to be replicated over a larger data set to be considered significant.

Competitive language was by far the most interesting metric. The assimilation hypothesis predicts that male players on matched avatars (MM) would use more competitive language than female players on matched avatars (FF), and that a mismatched avatar should decrease competitive language for men on female avatars (MF) but increase it for women on male avatars (FM). In practice, the sample showed exactly the opposite – a statistically significant correlation ($p = 0.0129$) between increased use of competitive language and female avatars, regardless of player gender. The highest counts were logged in the female matched condition (FF) and the lowest in female mismatched (FM), with male players falling somewhere in between. Male players logged a slightly higher but non-significant mean of competitive language in the mismatched condition (female avatar, MF). In short, the competitive language and emotional language metrics defy both folklinguistic stereotype and the assimilation hypothesis to a statistically significant degree. This suggests both hypotheses are insufficient, and another factor may be at play.

4.2.4 Gameplay context as a shaper of language

In the data, female players' use of emotional language features was significantly greater than that of male players ($p = 0.0128$). However, female players' use of emotional language increased rather than decreased with a mismatched avatar, creating even more stereotypically female language when they played a male avatar ($p = 0.0198$). The male-typical marker competitive language followed a similar but opposite pattern: female players' use of competitive language was significantly increased when their avatar was also female ($p = 0.0129$), creating more male-typical language in a matched condition. The shift can be seen in the paired samples of the female player who submitted logs in both the matched and mismatched condition. In short, both markers appear to rather sharply contradict the predictions of the assimilation hypothesis.

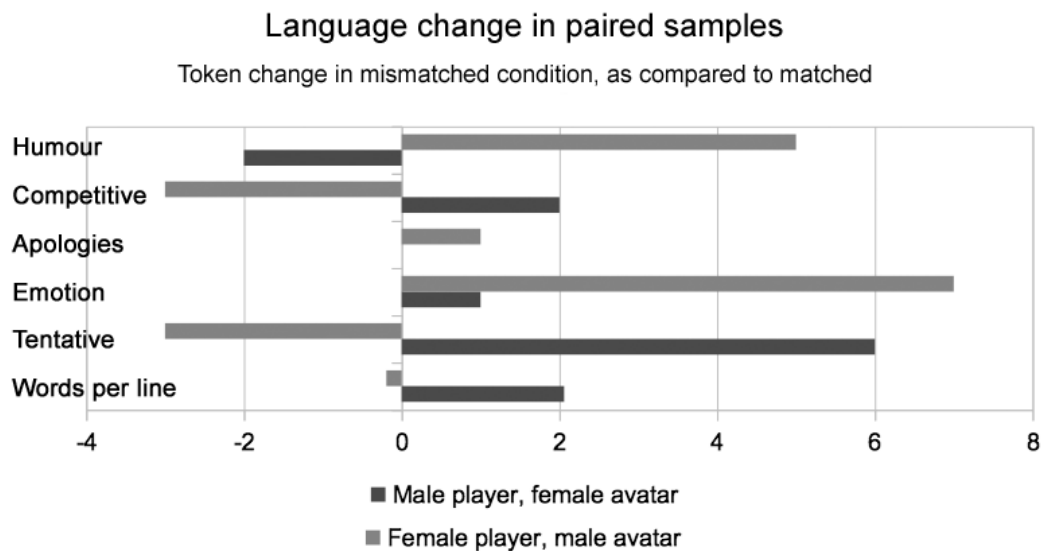


Table 6: Analysis of paired logs, showing how the token counts for various markers change in a mismatched condition, using the matched condition as a baseline. Words per line replicate the folklinguistic belief that women speak more than men do, as predicted by the assimilation hypothesis. Tentative language and humour also perform as predicted by the assimilation hypothesis, as do male tokens for emotion. Competitive language and emotional expression, however, defy expectations by shifting in the opposite direction: emotion only for the female player, competitive language for both.

A possible explanation is suggested by a close reading of the logs, in combination with the gameplay preference data reported by players and datamined from *WoW Armory*. In their demographic data, both female and male players reported interest in tactical gameplay (defined here as an interest in PvE and/or PvP content) at

approximately similar levels (2/3 of responders for each group). This would make the sample appear balanced. However, in the actual log excerpts cut for analysis, only the FF logs and one MF log contain tactical content, skewing the sample overall and the male (both player and avatar) samples in particular in favour of social rather than competitive interactions. Solutions for eliminating this sampling bias will be explored in the Discussion section; for now we simply review how gameplay context appears to shape gamer language.

Logs recorded in a tactical context register low emotional markers and high competitive markers: this correlates the player's language use with the gameplay context and self-identification as a competitive player, rather than with the player's or the avatar's gender. This is consistent with studies on gendered language in weblogs, which showed that use of gendered markers correlates more closely with the conventions appropriate to the genre of their weblog than the writer's actual gender: women writing informative logs use more masculine language than men writing personal journals (Herring & Paolillo 2006). Further, Bamman & al (2014) found that users in Twitter will employ gendered language not as predicted by their self-reported gender, but as predicted by the gender homophily of their social networks. Since studies on gamer culture suggest gamer identity is still construed as highly masculine (Shaw 2010), and *WoW* folklore holds that the playerbase is "about 2/3 male" (no formal statistics exist, though the number is frequently replicated on player forums and in community media), it is plausible that players identifying as competitive gamers may adjust their language in a similar way to webloggers, adopting the contextually preferred male-typical markers in a competitive (and arguably masculine) gameplay context, and using softer, female-typical markers in a social context.

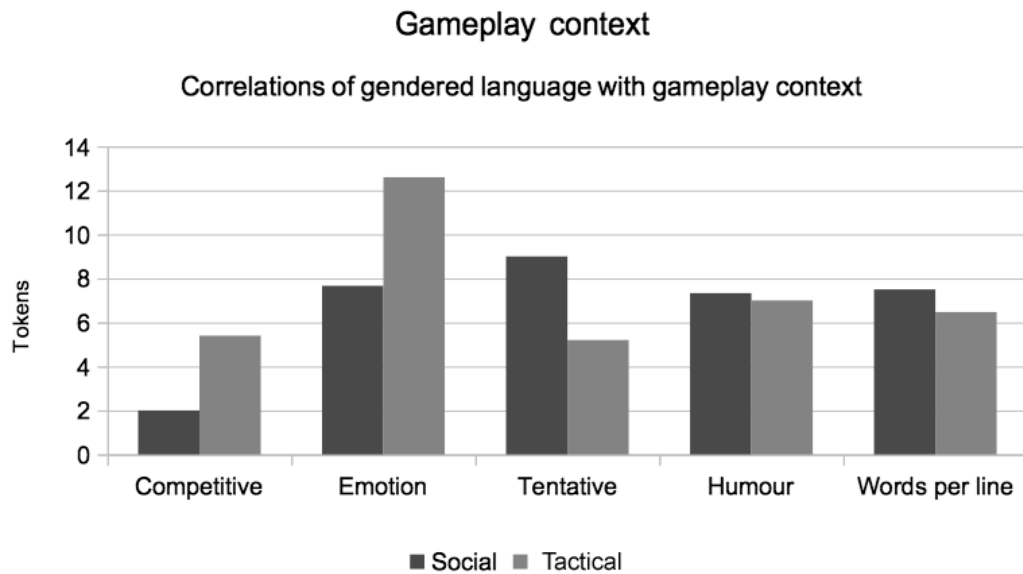


Table 7: Gendered language by gameplay context. Sorted by a player's self-reported focus on social activities (2 responders, 3 logs) or tactical activities (4 responders, 5 logs), tactical contexts prompt a statistically significant increase in competitiveness ($p = 0.0333$). Though emotional language and tentative language also respond to context, these shifts are not statistically significant. Of other observed metrics, humour and words per line seem stable regardless of context; apology and taboo are not shown due to low token counts overall.

This hypothesis finds some support in the data, which shows a statistically significant correlation between increased competitive language markers and a player's investment in competitive gameplay ($p = 0.0333$, investment determined as a combination of self-reported interest in tactical gameplay, as well as prestigious achievements in current competitive gameplay content listed in the avatar's Armory profile). No similar correlation was found for gameplay preferences and player gender, avatar gender, or player/avatar pairs. The relatively close correlations between language/gameplay context and between language/female players suggests that the gender correlation is incidental, and that the likelier shaper of linguistic choice is activity. Certainly this seems the more conservative estimate.

Emotional language also correlated with gameplay, but in a slightly more sophisticated way. Measured against player's preferred gameplay activities, neither the lexical nor composite measurements of emotional language showed a correlation. Emoticons, however, were significantly more prevalent ($p = 0.0179$) in the samples

of those players who also listed a competitive interest. This is not entirely surprising: the high speed interactions of competitive gameplay put a premium price on each keystroke, and the language of competitive gamers often reflects this, preferring social expressions like *:D* over *Awesome!* to free up valuable seconds for more optimised movement and gameplay. (This may also be a factor in why female verbosity in the matched condition was relatively low: logs recorded in a competitive context are necessarily shorter. This doesn't negate the finding, though – if anything, it suggests that if the sampling context had been consistent across the board, the change in verbosity for female players would have been more pronounced, and the result more significant.)

Overall, statistical testing replicates the findings of Palomares & Lee 2010 for tentative language, but not for apologies or emotional language. Emotional language, in fact, outright contradicts their findings, most likely due to variance in the sampling context – as the authors state, the shifts they observed were small, and likely to be buried by other factors conspiring to be more salient to the interaction than gender. This assumption is supported by the findings on competitive language and gameplay context: gamer identity would seem to overwrite gender identity, shown by competitive language mapping to context with the same approximate likelihood as it did to avatar gender in the FF sample.

4.2.5 Gendered use of parts of speech

The study was originally designed for use with the Palomares & Lee method; corpus methods were introduced after the sampling had already been done, to explore whether the study could be performed on a larger sample in a less labour intensive way. For now, any plans on doing so remain provisional. PoS-tagging with ARK-tagger produced usable raw data (and the manual emoticon count was replicated using it). However, statistical analysis found almost no correlations. Numerals, defined as both lexical and numeric expressions, did not show the expected correlation to masculinity for players or avatars. Pronouns showed only one significant correlation: the personal pronoun *you* in FF/FM pairs was shown to decrease on male avatars ($p = 0.0194$), corresponding to previous findings linking its use with female writers (Bamman & al. 2014, Argamon et al 2003). However, when so many other features are also indicated as gendered and yet not present, a lone

correlation seems more coincidental than indicative of meaningful variation in the data.

At least two factors may be contributing to the derth of results from PoS-analysis. First, the sample is tiny. While more qualitative manual coding can pull meaningful data from smaller samples via reading at a pragmatic and utterance level, corpus methods are conceived for use on significantly larger datasets, which are more likely to yield significant concentrations of structural units such as syntactic items and n-grams. In small samples, correlations found at a demographic level are easily ost due to the personal idiosyncrasies of outliers (such as a player who categorically refuses to use gendering pronouns, and so by Argamon's measures would read as male, or a very pronoun-prolific player, who by Bamman's reckoning would read as female). Secondly, second language speakers may simply not be viable for analysis by methods devised on first language sources. While neither study explicitly controlled to eliminate second-language speakers, they did target native speakers in a geographically appropriate area, and any second language segment in the sample was incidental. It is possible that functionally fluent second language English is still structurally distinctive in ways that render metrics calibrated for native speakers nonviable. For example, a Swedish player logged over twice the number of plural pronouns compared to other players – this could be (albeit tenuously and somewhat folklinguistically) argued as an artefact of the importance of negotiating consensus in Swedish culture.

A larger sample targeting native speakers may yield more useful results: for this sample syntactic analysis is of no particular utility, as it seems second language speakers simply don't conform to either a stereotypical or an assimilative model.

5 Discussion

The analysis shows the data to be uneven and only partially in line with expected hypotheses. Given the extremely small sample ($n = 6$ responders, submitting $n = 8$ logs, with $n = 2$ for paired samples of a player in matched and mismatched condition), the study can make no statistical claims of note: small samples are always more vulnerable to outliers, and it is entirely possible that a larger sample will either find correlations where this sample did not, or be unable to reproduce the

correlations found in this sample (particularly for emotional and competitive language, whose performance here is likely due to a sampling error). Several metrics shift in the expected direction, but even those shifts that were statistically significant are rendered provisional by sample size and bias. No significant conclusions on the presence or absence of assimilation can be drawn; nor is it expected of the pilot study. The purpose here was always to explore the viability of various methods for testing the assimilation hypothesis, and to establish a more refined method for a subsequent, representative study.

5.1 Refining the study design

This section examines the sources of error and variance in this study, and proposes an improved experiment design that will hopefully address the problems with the sample in this study.

5.1.1 Improving sample stability

The catchall way to improve statistical representativity of a sample is to increase sample size to the point where outliers no longer significantly impact the mean: this is simple with corpus methods, but with a human reading analysis directly adds to the time spent coding logs, and puts an outer limit on what can be considered a reasonable processing time. Since manual coding far outperformed corpus methods in terms of useful results from this study, it seems the more relevant method to allow for. This puts some pragmatic limits on how far sample size can be increased – but an n of ca. 40-120 responders (depending on sampling method) supplying 400-500 words each should provide a significantly more stable sample.

The sample's usefulness can also be increased by controlling for those factors that in this study were found to interfere with the tracked metrics: primarily gameplay context, sample length and the player's native language. The ambiguity introduced by gameplay context can be corrected in either of two ways: a larger sample to eliminate variance due to sampling or personal habits (the second can also be improved by using only paired samples), or a more standardised logging context, either communicated to responders *a priori* or selectively mined from submitted logs *post hoc*. The decision hinges largely on which is to be prioritised: the more

comprehensive overview of gamer parlance allowed by larger raw sample size, or the greater economy allowed by the smaller n in a restricted sampling situation.

Much of the ambiguity in the sample comes from the variety of speech contexts that arise in MMO gameplay – social, tactical and roleplay contexts all produced different profiles of gendered language markers (chiefly seen in competitive, tentative and emotional markers). This could be controlled for by limiting logging contexts to a particular sort of activity, ideally one as commonplace as possible to mitigate the impact of context-driven language and promote the salience of personal expression. Levelling and daily questing are activities undertaken by all players on a regular basis, including those whose investment in the game is primarily social. Targetting such low pressure contexts should eliminate the sampling bias of tactical gameplay being over/underrepresented, and help minimize the impact of strong gamer identification on use of gendering language. That the limitation is observed can readily be confirmed from logged chat: the relevant gameplay contexts are easily identifiable to a human reader familiar with the game.

Sample length in this study was ultimately defined by the necessity to rule out roleplayed content: this reduced the length of the shortest complete log from over 600 words to just over 300. This sort of data sparsity can be controlled for by limiting the context in which responders are asked to submit logs from. Sampling could also be limited to servers not labelled for roleplay (PvE and PvE, rather than RP-PvE and RP-PvE), making sure that all chat from a player is personal communication rather than mimetic drama. Compliance can be confirmed either on human read (roleplayed content is identifiable by its similarity to narrative prose) or in the stricter design from the player's demographic data (avatar name and server, here used for examining the player's ingame activities and responsibilities, would also allow confirming that the character's server is not labelled for roleplay). This should ensure that all logs submitted are sufficiently long to provide useful amounts of tokens for analysis. Focusing analysis on other server types may also improve the usefulness of the taboo language metric: roleplay-tagged servers are reputed to have a more polite community, which may account for the low token count found in this sample.

Some variation in the sample may also have been introduced by the responders being a randomly selected set of nationalities: analysed logs were controlled for player and avatar gender, but not for nationality. Since all responders were second language speakers, it is possible the idiosyncrasies of their native languages and/or cultural contexts influenced them: this seems especially likely with syntactic features, which in this sample correlated with previous research only on one metric, the use of *you* as indicating a female speaker. This can be controlled for by selecting as responders only native language speakers from a particular geographic area. For comparability with previous research, first-language speakers residing in the UK area may be the most fruitful demographic to target.

It's worth noting that player nationality, much like player gender, is accessible by self-reporting only. While gameplay-context can be deduced from logged text and avatar gender can be verified from *WoW Armory*, player gender and nationality must be supplied by the responder, and accepted at face value.

Finally, ideally all responders would submit a sample in both the matched and mismatched condition. This would allow the samples to be analysed as paired, each acting as their own control. Combined with measures to standardise the logging context, this would eliminate most of the sources of variance that here complicated analysis.

5.2 Refining the analysis

The female-typical metrics used by Palomares & Lee (2009) and the masculine typical metrics derived from Coates (1986, 2003) and McEnery (2005) performed in comparable ways. Both categories contained a metric that produced insufficient tokens for analysis – apologies for feminine-typical and taboo for masculine-typical metrics – and as such, little can be said on them. Both categories also produced workable data: emotional and tentative language in the female-typical metrics, and competitive language in the masculine metrics. Failure to account for gameplay context in sampling makes emotional language and competitive language in this sample useless as indicators of assimilation – rather, assimilation is found to a social or competitive gamer identity dependent on the gameplay context, and the presence of emotional or competitive markers corresponds to those. Humour conformed to the expected patterns of assimilation, but not at any level of significance. This leaves

only tentative language – and on this, an apparent pattern of assimilation seems to occur, with male players on female avatars increasing their use of tentative language. Refinement to the method here would entail two things. First, assigning coding of the data to condition-blind coders to rule out perception bias on the coder's part shaping token counts; and secondly, analysis of the coded data by ANOVA rather than simple T-testing, to replicate the method of Palomares & Lee in full.

5.2.1 Coding ambiguities for future consideration

Manual coding guidelines were mostly sufficient, but still produced some ambiguous tokens that for the purposes of this study were either included into the most relevant variables or left out. A larger sample may provide more evidence to support or contradict those choices – and so, the ambiguous cases are recorded, for later reconsideration.

Emotional expressions presented a variety of borderline cases. Gameplay chat uses several nonstandard linguistic practices for encoding the nonlexical and nonverbal cues of face-to-face interactions. The logs included multiple cases of these: emoticons (*^*, *:D*, *|o/*), interjections (*awww*, *eep*, *nnnnnrgh*), token emotes (**gigglesnort**, */cry*) and button mashing (*dfjgkhdfkjshk*). All suggest the presence of an emotional state, but do they constitute reference to emotion in the linguistic sense?

A meta-analysis of studies on CMC suggests that the primary function of emoticons appears to be as emblematic indicators for emotions that face-to-face would be obvious from nonverbal communication, and that no static split has been found between genders in how emoticons are used (Derks, Fischer & Bos 2008). Some corpus studies on emoticons have found that different emoticons are used in gender-preferential ways, positivity and solidarity for women and sarcasm for men (Wolf 2000) – the division does not, however, manifest in all online environments.

Ultimately the choice to include emoticons as emotional reference was made based on recent Twitter studies, in which emoticons were found to be preferred by female users (Bamman et al 2014). As such, they were tallied as a form of emotional language, and ultimately rolled into the composite measure.

Internet acronyms (*afaik*, *iirc*, *ffs*, *lol*) were parsed as the phrase they represent and included in their respective categories. Tentative language (*as far as I know*, *if I*

recall correctly) and taboo language (*for fucks sake*) were self-evident categorisations. *Lol* (and its *WoW*-slang counterpart, *kek*) appeared in two forms, as an interjected emblem of laughter, but also as a noun in phrases like *lolworthy* and *for the lulz*, and a verb in its own right in phrases like *I lol'd*. In both cases it was interpreted as an indicator of emotion, and coded as emotional language.

Buttonmashing – nonlexical strings of characters produced by mashing the keyboard – was considered for emotional language, but not included. In this sample buttonmashing was logged by a competitive female player, reconnecting for the final time after a series of disconnects killed her team and failed the timed challenge they had been pursuing. The violence implicit in an incoherent, buttonmashed string of alphabet soup is reminiscent of the visceral emotion of howling in frustration, or the pictograms scribbled into the speech bubbles of flailing cartoon characters screaming ultimate profanity. This suggests two viable codings, either as emotional language (an expression of rage) or as taboo language (parallel to nonswears such as *fudge*). This particular case of buttonmashing should perhaps be parsed as taboo language – it fits McEnery's definition of a nonswear by filling the syntactic place and semantic function of profanity with something less off-colour, and fits his finding that such phrasing is preferred by women. Though strong emotion is also implied, it clearly isn't what Lakoff would have called female-appropriate sentiment: “women can complain, but only a man can bellow in rage” (2004: 45), which would also support coding the token as taboo rather than emotion. Either way, a single token is insufficient material to draw such elaborate conclusions from, and as such, the token was left uncategorised, pending reconsideration in a broader sample.

5.2.2 Testing male-typical metrics

The lack of male-typical metrics to complement the female-typical ones in the study of assimilation was noted by Palomares and Lee (2009), and addressed in this study with a set drawn from quantitative research by Coates (1986, 2003) and McEnery (2005). For now the set must be considered a work in progress: although the coding guidelines seem comparable to those of the female-typical set, problems with the sample made it impossible to test them in detail. For taboo language, this meant an insufficient token count for analysis (much as was the case with the female-typical metric, apologies). For competitive language, the issue was with sampling, which led to an asymmetry between the contexts the logs were recorded in. In this study

competitive language correlates more strongly with gameplay context than with gender, and the details of this correlation cannot be reliably explored, as female players were logged primarily in competitive context, and male players primarily in a social one. It's worth noting that does not necessarily rule out the assimilation of gendered cues – just that the assimilation may be to the masculine gamer stereotype, rather than to the visually cued avatar. Nevertheless, it seems prudent to test the male-typical metrics on a less ambiguous sample for a better estimate of their validity.

While Coates (2003) observed an absence of narratives of victory in women's speech, Shaw (2010) notes that gamer identity is indexed strongly through straight white maleness, even among those not actually belonging to that demographic, and argues that for dedicated gamers the traditionally male narratives of competition, struggle, and the occasional defeat in pursuit of epic win are central communal experiences. Fluency in communal narratives falls under the umbrella of camaraderie-fostering behaviours expected of women (Herring & Martinson 2004), which seems to make competitive language a natural part of a female gamer's register. This mirrors Bamman et al's (2014) finding that gender homophily in social networks is a better predictor of gendered language use among its members than the gender of an individual speaker, and also fits the data examined in this study. In addition to submitting logs from a competitive gameplay situation, players who rated high in competitive expression also had significantly more competitive gameplay achievements listed on their Armory profiles. By communal standards, such achievements are considered indicative of a skilled and dedicated player.

In raw terms of producing countable tokens, observable patterns and a significant (even if unexpected) correlation, the proposed metrics for male-typical language seem worth applying to a larger sample. Even if competitive language proves unsatisfactory at indexing gender in a gaming setting, the correlation between competitive attitudes and committed gameplay remains interesting, and may have relevance to how games come to be construed as male environments.

5.3 Implications

Previous cognitive research has pointed to several implications of avatar assimilation outside mere academic interest. Firstly, that the virtual bodies we assume have real power to shape our perceptions of ourselves, and through those our immediate

behaviour towards others – and secondly, that these changes persist at least a short while outside the virtual world, making us act more confidently after inhabiting tall or beautiful avatars (Yee 2007) or view others with less racist assumption after assimilation to a racialised appearance (Peck & al 2013). Walking a mile in virtual shoes does seem to breed connection with the identity we adopt, and that connection can be used to increase empathy and understanding.

The same process, however, can also work counter-productively. Social psychology has found that exposure to the strongly sex-typed characters that are common in videogames led men to be more accepting of sexual harassment, an effect that was also extended to rape-myth acceptance (i.e., the belief that rape victims share in the blame for being targeted) on prolonged exposure (Dill, Brown & Collins 2007). Another study found that women assigned a sexualised avatar in a virtual reality experience reported more body-related thoughts after the experiment than women assigned a non-sexualised avatar; if participants also saw their own face on the avatar through a virtual mirror, they expressed more rape myth acceptance than those who did not (Fox, Bailenson & Collins 2013). As an identity, 'gamer' is still predominantly indexed as male, white, technically savvy and heterosexual (Shaw 2010), a hegemony reinforced on the one hand by the AAA-list single-player game's ubiquitous white male protagonist, and on the other by the hypersexualised, male-gaze rendered avatars of MMO's, including *WoW*. While video games are far from the only form of popular culture playing to that stereotype, their participatory nature and the weight of dozens, even hundreds of hours sunk into a single game may render the specifics of video game representations much more significant to players than those of the latest Hollywood blockbuster, once watched and quickly forgotten. This puts a rather disturbing spin on video gaming's persistent problem with sexism and harassment. All major online game services now include a function for reporting or at least muting inappropriate messages. Players will frequently complement these functions with peer surveillance of their own (Collister 2014). A study of 300 *World of Warcraft* players found that just over half of male players and three quarters of female players had witnessed sexism against other players; 11.6% of males and 63.6% of females had experienced it personally (Brehm 2013). Studies on gender-swapping have without fail found that avoiding harassment is one of the major motivations for female players choosing male avatars (Hussain 2008; Lou et al

2013). The GamerGate debacle and the continuing harassment of developer Zoe Quinn (Romano 2014) is just one incident in a decade of harassment and threats against female designers and feminist games scholars (Chess & Shaw 2015), serving to show that even with women accounting for 44% of the US videogamers and underage boys clocking in at only 15% (ESA 2015), some still regard the gaming as a boys' club where women are welcome only as objects, and female developers can only succeed by trading on their sexuality.

If avatar assimilation can be shown to be factual rather than hypothetical, game designers can no longer claim the choices they make in visual or animation design are purely aesthetic, without other implications. If we truly adopt and experience virtual bodies in ways similar to our physical body, then the values, affordances and sympathies imprinted on avatars, and by extension the player, are not incidental. This has already been accepted in terms of player preference, as seen in news like *WoW*'s sister game *Overwatch* redesigning a sexualised female animation in response to player complaints (McWherton 2016). Developers taking an interest in the affordances and behaviours avatars may impart on players has yet to happen, though – unless perhaps in the indie survivalist game *Rust*, which recently assigned race and gender randomly and permanently to all player avatars, calling this the more realistic option (Newman 2016). For video gaming to truly come of age as an art and as a medium, it will have to accept that art is not created in a vacuum, or limited to the context in which it is enjoyed. The structures that a culture presents will shape the structures that its members reproduce.

This connection of representation and reproduction is exactly what I've attempted to show in this study, by drawing a link between gendered avatars and the gendered language produced by the players behind them. Linguists have at their disposal powerful means of connecting language to social structures, exposing meanings underlying the surface level of words and sentences. The happy obsolescence of some of Lakoff's original case examples – *lady* falling out of use as a relevant qualifier from *doctor* or *scientist*, and *professional* becoming genuinely disgendered rather than a euphemism for prostitute – is a case example of how language can serve as a diagnostic to expose underlying internalised values, and how political action to change those values will in time also feed back to change the language.

6 Conclusions

The design of the study was to establish and test metrics for examining avatar assimilation along both feminine and masculine paradigms in a natural virtual world setting. I hoped to investigate a number of things. First, whether the visual gender of a videogame avatar will skew the player's linguistic choices in favour of the gender mapped by the avatar. Second, to identify features other than avatar gender that may influence player's gender presentation online. Third, to test a set of male-typical language features that might be used as a complement to the established female-typical metrics.

The pilot study cannot claim to have accomplished either of these things, as sampling was demonstrably skewed and significance cannot be presumed with this sample size, even if it had been unbiased. Limited evidence of avatar assimilation was found in the raw data of tentative language for male players and humour for all players, as well as in verbosity for all players, but neither correlation was statistically significant. A strong correlation was found between some competitive language (a masculine marker) and competitive gameplay activities, suggesting that logging context will significantly impact the specifics of the language logged. The male-typical metrics performed in a comparable way to the established female-typical metrics, and seem worth developing further as a diagnostic for stereotypically masculine language – however to establish and if necessary improve the validity of selected metrics, additional testing should be performed on a more conventional dataset.

Overall, with properly refined sampling procedures, a larger sample and a more thoroughly tested metric for male-typical language, more reliable results seem obtainable.

7 References

7.1 Bibliography

Argamon, S., Koppel, M., Fine, J. & Shimoni, A. R. (2003) Gender, genre and writing style in formal written texts. *Interdisciplinary Journal for the Study of Discourse*. Volume 23, Issue 3, Pages 321–346

- Bem, D. (1972). Self perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology (Vol. 6)*. New York: Academic Press.
- Butler, J. (1990) *Gender Trouble*. New York: Routledge.
- Chess, S. & Shaw, A. (2015) A Conspiracy of Fishes, or, How We Learned to Stop Worrying About #GamerGate and Embrace Hegemonic Masculinity. *Journal of Broadcasting & Electronic Media*, 59:1, 208-220.
- Coates, J. (1986) *Women, Men and Language*. Longman House: Harlow.
- Coates, J. (2003) *Men Talk*. Stories in the making of masculinities. Oxford & al.: Blackwell publishing.
- Collister, Lauren B. 2014. Surveillance and Community: Language Policing and Empowerment in a World of Warcraft Guild. *Surveillance & Society* 12(3): 337-348.
- Derks, D., Fischer, A. H. & Bos A. E. R (2008) The role of emotion in computer-mediated communication: A review. *Computers in Human Behaviour*, 24, 766-785.
- Dill, K. E., Brown, B. P. & Collins, M. A. (2007) Effects of exposure to sex-stereotyped video game characters on tolerance of sexual harassment. *Journal of Experimental Social Psychology* 44 (2008) 1402–1408.
- Entertainment Software Association (2014). *Industry Facts*. Read 18.02. 2016. <http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf>
- Foucault, M. (2010) *Seksuaalisuuden historia*. Suom. Kaisa Sivenius. Helsinki: Gaudeamus.
- Frank, M. & Gilovich, T. (1988). The dark side of self and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, 54, 74-85.
- Fox, J; Bailenson, J.N., Tricase L (2013) The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behaviour*, 2013 29: 3, 930-938.
- Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J. & Smith, N. A. (2011) Part-of-Speech Tagging

- for Twitter: Annotation, Features, and Experiments. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, companion volume, Portland, OR, June 2011.
- Hall, K. (2004) *Language and Marginalised Places*. In Lakoff, R. ed. Bucholtz, M., 171-178.
- Herring, S. (2004) *Computer-Mediated Communication and Women's Place*. In Lakoff, R. ed. Bucholtz, M., 216-222.
- Herring, S. & Martinson, A. (2004). Assessing Gender Authenticity in Computer-Mediated Language Use. Evidence From an Identity Game. *Journal of Language and Social Psychology*, 2004, 23: 4, 424-446.
- Hussain, Z. & Griffiths, M. D. (2008): Gender Swapping and Socializing in Cyberspace: An Exploratory Study. *CyberPsychology & Behaviour*, 2008 11: 1, 47-53.
- International Software Federation of Europe, (2012). *Videogames in Europe: Consumer Study. European Summary Report*. Read 03.02.2016.
http://www.isfe.eu/sites/isfe.eu/files/attachments/euro_summary_-_isfe_consumer_study.pdf
- Johnson, R. & Downing, L. (1979). Deindividuation and valence of cues: Effects on prosocial and antisocial behaviour. *Journal of Personality and Social Psychology*, 37, 1532-1538.
- Lakoff, R. (2004). *Language and Woman's Place*. Text and Commentaries. ed. Bucholtz, M.. New York: Oxford University Press.
- McGonnell-Ginet & al. (1980) *Women and Language in Literature and Society*. New York: Praeger.
- McWherton, M. (2016) Blizzard is removing a sexualized pose from Overwatch, citing player feedback. *Polygon*. Read 14.4.2016.
<http://www.polygon.com/2016/3/28/11321138/overwatch-tracer-pose-removal>
- Merola, N., Penas, J., & Hancock, J. (2006). *Avatar color and social identity effects: On attitudes and group dynamics in virtual realities*. Paper presented at the ICA 2006, Dresden, Germany.

- O'Barr, W. & Atkins, B. K. (1980). 'Women's language' or 'powerless language?', In McGonnell-Ginet & al. pp. 93-110.
- Newman, G. (2016) Why my videogame chooses your character's race and gender for you. *The Guardian*. Read 14.4.2016.
<http://www.theguardian.com/commentisfree/2016/apr/13/videogame-chooses-character-race-gender-rust>
- Newman, M., Groom, C., Handelman, L. & Pennebaker J. (2008) Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes* 45: 211–236.
- Palomares, N. A. & Lee, E. (2009) Virtual Gender Identity: The Linguistic Assimilation to Gendered Avatars in Computer-Mediated Communication. *Journal of Language and Social Psychology*, 2010 29: 1, 5-23.
- Peck, T. C., Seinfeld, S., Aglioti, S. M., Slater, M. (2013) Putting yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias. *Consciousness and Cognition*, 22: 3, 779-787.
- Romano, A. (2014, August 20). The sexist crusade to destroy game developer Zoe Quinn. *Daily Dot*. Read 18.2.2016. <http://www.dailydot.com/geek/zoe-quinn-depression-quest-gaming-sex-scandal/>
- Statista (2015). *World of Warcraft Subscribers by Quarter*. Read 03.02.2016.
<http://www.statista.com/statistics/276601/number-of-world-of-warcraft-subscribers-by-quarter/>
- Warcraft Armory*. Read 02.03.2015. <http://eu.battle.net/wow/en/community/>
- Williams, D., Consalvo, M., Caplan, S., & Yee N. (2009) Looking for Gender: Gender Roles and Behaviors Among Online Gamers. *Journal of Communication*, 59.4, 700-725.
- Yee, N. (2007) *The Proteus Effect: Behavioural Modification via Transformations of Digital Self-Representation*. Read 02.03.2015.
http://www.nickyee.com/pubs/Dissertation_Nick_Yee.pdf
- Yee, N., Ducheneaut, N., Yao, M., & Nelson, L. (2011) Do Men Heal More When In Drag? Conflicting Identity Cues Between User and Avatar. *Proceedings of CHI 2011*, 773-776.

7.2 Secondary Literature

Daedalus Project. Read 02.03.2015. <http://www.nickyee.com/daedalus/>

Poles, K., Ijsselstein, W.A. & Kort, Y. (2015). World of Warcraft, the aftermath: How game elements transfer into perceptions, associations and (day)dreams in the everyday life of massively multiplayer online role-playing game players. *New Media & Society*. 2015 17:7, p. 1137-1153.

7.3 Software

Carnegie-Mellon ARK Tagger (2013, version 0.3) [Mac OS X 10.11.13], documented in Gimpel & al (2011). Available from <http://www.cs.cmu.edu/~ark/TweetNLP/>

Anthony, L. (2014). AntConc (Version 3.4.3) [Mac OS X 10.11.13]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

Sinclair S. & Rockwell, G. (2016). Voyant Tools (version 1.0) [Mac OS X 10.11.13]. Available from <http://www.voyant-tools.org/>