UNIVERSITY OF
EASTERN FINLAND

The Use of Taboo Language in a Corpus of Chat Messages of *Defence of the Ancients 2*

Mikko Halonen
English language and Culture
University of Eastern Finland
Philosophical Faculty
School of Humanities
4.3.2024

Itä-Suomen yliopisto, filosofinen tiedekunta
Humanistinen osasto
Englannin kieli ja kulttuuri
Halonen, Mikko: The Use of Taboo Language in a Corpus of Chat Messages of Defence of the Ancients 2
Opinnäytetutkielma, 76 sivua
Tutkielman ohjaajat: professori Mikko Laitinen
Maaliskuu 2024

Asiasanat: DotA 2, epäkohteliaisuus, face, kasvot, verkkopelit, kiroilu, kirosanat, frekvenssianalyysi, toksisuus, toksinen kieli, käyttäjälähtöinen aineisto, verbaali häirintä

DotA 2 on Valve-yhtiön julkaisema ja ylläpitämä ilmainen, verkko- ja joukkuepohjainen strategiapeli, jota miljoonat pelaajat pelaavat päivittäin. Tämän tutkimuksen tavoitteena oli tarkastella tabukielen käytön laajuutta 25 useimmin käytetyn kirosanan avulla DotA 2 -verkkopelissä.

Tutkimuksen teoreettinen viitekehys rakentuu olennaisesti kohteliaisuuden ja epäkohteliaisuuden käsitteen face (suom. kasvot) ja face-threatening-acts (suom. kasvoja-uhkaavat-teot) varaan. Tutkimuksen teoreettisessa viitekehyksessä hyödynnetään myös laajalti kiroiluun, toksiseen kieleen ja verkkopeleihin liittyvää tutkimusta, ja ns. tabukielen käsitteistöä käytetään aineiston tulkinnassa suurissa määrin.

Tutkimuksen aineistona toimi verkossa vapaasti saatavilla oleva Dota 2 Matches datapaketti. Data sisältää viestejä neljän vuoden ajalta, joita pelaajat lähettävät toisilleen pelin aikana. Viestit poimittiin korpukseen ja analysoitiin AntConc-korpustyökalulla (Anthony 2023). Standardienglannin vertailuaineistona tutkimuksessa hyödynnettiin Contemporary Corpus of American English (COCA) ja British National Corpus (BNC) -korpuksien suullisia alakorpuksia. Suullisten alakorpusten genrejä ei valikoitu tarkemmin tutkimuksen laajuuden vuoksi, minkä vuoksi vertailuaineistot sisältävät myös esimerkiksi suullisia uutistekstejä. Ensin kymmenen

yleisimmän lekseemin frekvenssit poimittiin korpuksesta, minkä jälkeen 25 yleisintä tabukielen lekseemiä poimittiin omaksi vertailuaineistokseen standardienglannin korpusten vertailun tueksi. Tutkimuksessa hyödynnettiin frekvenssianalyysiä, mitä tuettiin laadullisella konkordanssianalyysillä.

Tutkimuksessa havaittiin, että DotA 2 -verkkopelin pelaajat hyödyntävät runsaasti akronyymejä ja lyhennelmiä kuten *wp*, *gg*, *ez* ja *lol*, joista *lol* onkin frekvenssiltään suurin. Havaittiin, että kymmenen yleisimmän lekseemin joukossa eri standardinenglanninvastaiset kirjoitusasut ovat yleisiä ja erityisesti persoonapronominien kirjotusasujen frekvenssit ja eri muodot kuten /u/ ja /you/ nauttivat runsasta käyttöä. Todettiin, että pelaajat suosivat nopeaa ja tehokasta viestintätyyliä, minkä lisäksi aineistossa on viitteitä selvästä pragmaattisesta kohteliaisuusnormista wp-akronyymin käytön vuoksi. Aineistolle tyypillisimmät kirosanat *fuck*, *fucking* ja *shit* koostivat valtaosan tuloksista ja yleisimmiksi tabukielen kategorioiksi viestikorpuksessa paikannettiin seksuaaliset viittaukset ja loukkaavat nimet ja herjat. Todettiin, että yleisimmät kasvoja haastavat tai uhmaavat teot materialisoituvat tutkimuksessa herjoihin kuten *retard* ja *idiot*. Vertailussa standardienglannin aineistoihin havaittiin, että kirosanojen ja tabukielen frekvenssi viestikorpuksessa on yleisimpien lekseemien kohdalla liki 400 kertaa korkeampi kuin COCA ja BNC-korpuksissa. Vertailussa viestikorpuksen ja BNC-korpuksen kanssa havaittiin, että korpukset kuitenkin jakavat kolmen yleisimmän lekseemin frekvenssin, kun verratessa COCA-korpukseen vastaavaa yhtäläisyyttä ei havaittu.

Tutkimuksen perusteella todettiin, että virtuaalisen maailman aineistot poikkeavat olennaisesti standardienglannin kielestä. Tämän tutkimuksen aineisto sisälsi moninkertaisen määrän kaikkia vertailtuja lekseemejä ja vertailuaineistoista vain BNC-korpuksesta voitiin havaita lieviä vastaavuuksia. Tutkimustuloksista käy ilmi, että puhujien valmius tabukieleen anonyymissä verkkoympäristössä on selvästi poikkeava standardienglannista ja jatkotutkimus on selvästi tarpeen.

University of Eastern Finland, Philosophical Faculty
School of Humanities
English Language and Culture
Halonen, Mikko: The Use of Taboo Language in a Corpus of Chat Messages of Defence of the Ancients 2
Thesis, 76 pages
Supervisors: Professor Mikko Laitinen
March 2024

Keywords: DotA 2, impoliteness, face, online gaming, swearing, swear words, taboo language, frequency analysis, toxic language, toxicity, user-generated data, verbal harassment

DotA 2 is a freely available online video game published and distributed by the Valve Corporation. The game is played by millions of players daily and is based on cooperative team-based gameplay. The aim of this paper was to examine the prevalence of taboo language by examining the 25 most frequent types of curse words used by players of DotA 2 in comparison with standard English data.

The theoretical framework for the study is constructed around the notions of politeness and concepts such as face and face-threatening-acts, however research dealing with cursing, toxic language online, and online gaming heavily contributed to the theoretical framework of the paper.

The data for the study was derived from a freely available online data set Dota 2 Matches, which contains chat messages sent by players to each other during matches covering a 4-year period. The messages were collected and analysed using the AntConc –corpus analysis tool (Anthony 2023). Comparisons between standard language were facilitated by the spoken subcorpora of the Contemporary Corpus of American English (COCA), and the British National Corpus (BNC). The subgenres of the corpora were not further limited due to the scope of the paper. First, the frequencies of the 10 most frequent items were surveyed, after which the frequencies of the 25

most frequent curse words were collected and compared to the standard English corpora. The present paper's analysis relies on frequency analysis, supplemented by qualitative concordance analysis.

It was noted that players of DotA 2 favour frequently employ acronyms such as *wp*, *gg*, *ez*, and *lol*, with lol exhibiting the highest frequency. It was found that within the 10 most frequent items, non-standard spelling forms of pronouns such as /u/ and standard forms such as /you/ enjoy abundant use. It was found that players employ efficient and swift communication styles, and that the use of such acronyms such as *wp* may hint at a pragmatic norm in the medium. *Fuck*, *fucking*, and *shit* were highlighted as the most frequent types of taboo vocabulary in the chat corpus, and the most frequent categories of taboo words were comprised of sexual references, and pejoratives. The results indicate that the most common face-threatening-acts may be realised by pejoratives such as retard and idiot. In comparison with standard English data, it was found that players in DotA 2 employ taboo language nearly 400 times more frequently than speaker in the BNC, or the COCA. However, it was found that the BNC exhibits the most likeness with the chat corpus in the three most frequent types of taboo vocabulary, i.e., *fuck*, *fucking*, and *shit*. Conversely, the COCA did not exhibit comparable similarities.

It was concluded that user-generated linguistic data from online spaces may drastically differ from comparable data sets of standard English. The data of the present paper contains remarkably higher frequencies of taboo language compared to data from standard English corpora such as the BNC or the COCA. Although neither of the compared standard corpora displayed nearly similar frequencies, it was found that the BNC may bear the highest similarity in terms of the three most frequent items in the chat corpus. The results of the paper clearly indicate that speakers behave in drastically different ways in online environments, and their readiness to deploy taboo language may be much higher when shrouded in anonymity. The present paper highlights a glaring need for further study into the language of online gaming.

Contents

1 Introduction

Abusive language has been recognised widely in the study of online discourse, and its detrimental effects on individuals' wellbeing and enjoyment of and within online gaming has been widely documented by previous literature. This thesis examines the usage of taboo words as vehicles of impoliteness in a corpus of chat messages from the *massively online battle arena* (MOBA) online video game Defence of the Ancients 2 (DOTA 2). This study will attempt to examine verbal toxicity in online gaming through a specific framework of the MOBA title DotA 2 by utilizing a corpus of text chat messages produced by players during gameplay in ranked matches. The study will approach the topic through a theoretical framework combining the notions of impoliteness and face-threatening attacks, i.e., how speakers use words to deconstruct and challenge their co-locutors' pragmatic self-image in relation to oneself and others. The aim of this thesis, then, is to assess the level of abusive language in the gaming context. Additionally, a general overview of 'taboo' language will be produced to facilitate comparisons between different corpora. The results of the study are further compared to major corpora of the English language like the BNC and COCA, in order to provide comparisons, and contextualise the distributions within. By analysing the quantitative dimensions of the taboo language that lay in the medium, my aim is to provide a snapshot of the discourse inside the gaming environment with a specific focus on profanity and verbal toxicity as outlined in the literature. In this study, several research questions are put forward:

1. What are the most frequent items in the corpus of chat messages, and how do they characterise the discourse?
2. What are the frequencies of the 25 most frequent taboo words in the chat corpus compared to two standard English language corpora, the British National Corpus, and the Corpus of Contemporary American English?
3. How can notions of politeness and face be applied to the data?

Research into online gaming, and specifically into MOBAs like DotA has seen relatively few papers from the perspective of linguistics, though different papers have been published delving into various discourse features of online video games. Studies on the phenomenon of *toxicity* i.e., abusive verbal and non-verbal in-game behaviour, such as the abuse of in-game mechanics to distort match outcomes, have been plenty, and authors like De Mesquita Neto and Becker (2018) and Lee et al. (2022), for example have investigated the topic in the context of another popular MOBA title "League of Legends" with a focus on conversational patterns and their relationship to toxicity. Lee et al. (2022) examines the player perception of toxicity and its relationship to engagement in online gaming in the context of Dota 2.

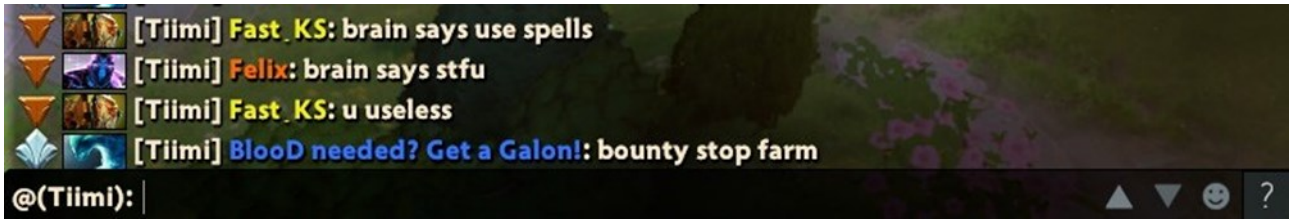

Figure 1. An illustrative screenshot of a chat interaction in DotA 2.

Aside from the lack of papers regarding the linguistics inside the medium, papers on the topic have dealt with the analytical and artificial intelligence related aspects of the medium like predicting match outcomes (Kodirjon and Anh Huy 2021). However, a handful of authors (Kwak and Blackburn 2015) have examined the linguistic elements of toxicity through the lens of toxic verbal behaviour. Although research into the linguistic activities inside the medium have been scarce, insight into toxic language is certainly warranted given the negative effects abusive language may have on an individual. Fox and Tang (2017: 1298) indicate that verbal and sexual harassment directed at different groups of people, like women in gaming environments, may have a number of negative consequences like withdrawal from gaming situations and rumination of the abuse far after the gaming context where the abuse took place. Verbal harassment and online misdemeanour also lower players' enjoyment of a game (Teng et al. 2008), and as such it is important to assess the spread of toxic verbal behaviour.

Study into the linguistic activity taking place in these online environments is clearly marked by a need for further research, and awareness of toxic language in online gaming should be highlighted (Mattinen and Macey 2018: 7, Lee et al. 2022: 4). Research into the linguistic devices used to perpetrate verbal abuse provide key information to prospective gamers and enthusiasts, but also to help inform guardians', parents', and educators' views on online gaming to better prepare vulnerable individuals. In addition, any commercial actor interested in maintaining an inclusive and welcoming gaming environment will surely find any characterisations and patterns of toxic language patterns helpful. Researchers looking at swearing, like Jay (2009: 157), note that the very topic of swearing, and what motivates it has not received adequate attention from the academic research community, and so this paper serves to extend the research of swearing into online spaces by providing unique perspective into a less studied context of language use, and as Culpeper (2013: 3) additionally underscores, the topic of impoliteness is viewed somewhat negatively by the academic community as 'dirty', and due to its perceived one-dimensionality, not worth analysing.

As a core linguistic feature examined in this paper relates to profanities, insults and general 'bad language', this paper serves as a vehicle to expand the branch of research dealing with non-standard language. A number of terms such as 'bad language' (Coats 2021), 'taboo language' (Jay 2009: 154), and 'socially opprobrious language' (Sulpizio et al. 2019: 84) have been previously used to address the phenomenon, and though these terms emphasise different aspects of what might be described as taboo, or inappropriate language, this paper adopts the more general descriptor 'taboo language' in line with Jay (2009). In contrast, the term *toxic language* in sections 2.1-2.3 relates to abusive language more generally within the virtual environments under discussion, and not 'cursing' or taboo language as a linguistic feature. The issue of terminology regarding taboo language will be addressed in more detail in section 3.1.

For the reasons outlined here, a study into the impolite language use in a game like DotA is certainly justified when considering the relative lack of academic interest in the subject, while still contributing directly to the study on in-game verbal toxicity and the study of the linguistics regarding taboo words, and impoliteness. This paper provides a particularly novel perspective on

impoliteness within computer mediated communication (CMC) studies, as, though papers on impoliteness in CMC contexts are plenty, few examine a gaming environment and speaker groups as the present paper.

First, a brief theoretical framework will be laid out for the reader providing a background for the present paper regarding, first, the conceptualisation of toxicity more broadly in the context of CMC and narrowing down to gaming spaces. Second, an examination of previous research into the gaming environment and a number of previous studies will be presented, and third, the literature section of this paper will cover various impoliteness theories, as well as the concepts of *face* and rapport management, relational politeness, and, finally a comprehensive summary of previous research on taboo language, as well as a definition for the phenomenon will be laid out. Subsequently, a brief description of the methods and the data used in the study will be outlined, and finally concluding on the results of the study and the discussion and conclusions. Additionally, directions for future research, as well as the limitations and shortcomings of the present study will be explored briefly.

2 Background and theoretical framework

2.1 'Toxicity' in online gaming

Previous literature provides several definitions and perspectives for *toxicity* and *toxic behaviour,* however the very concept proves an ambiguous phenomenon to disentangle in a singular definition. Toxic behaviours in gaming may be conceptualised more broadly, relating to extralinguistic and aspects particular to the individual virtual environment such as account theft, resource hogging, bullying through targeted harassment, profanity or cheating via specialised computer software or code (Teng et al. 2012: 351), sexual and general harassment, such as rape threats and sexually charged comments towards female players (Tang and Fox 2016, Fox and

Tang 2017), trash talking and cyberbullying (Ballard and Welch 2017), and racial prejudice or harassment, such as racial profiling and the deployment of racial epithets (Ortiz 2019). Related conceptualisations of reprehensible or harmful online discourse have touched on the notions of *flaming*, the insulting of the co-locutor and general "bad affect" derived from such discourse (Jane 2015: 66). Jane (2015: 66) provides further delineations for abrasive modes of online discourse, highlighting the concept of *trolling, "*the posting of deliberately inflammatory [...] material with the aim of provoking [...]", and *cyberbullying,* the targeted abusive language occurring mainly within educational settings among youths*.* The range of objectionable behaviours in gaming, however, range from *trash talking* to other forms of objectionable language (Tang and Fox 2016: 514).

More general definitions of toxic behaviours incorporate the idea of norm violation in which an act becomes 'toxic' by means of violating the "generally accepted norms" as a breach of norm in the broader discourse (Teng et al. 2012: 343)*.* Indeed, player experiences indicate a set of shared norms delineating appropriate behaviour exemplified by the boundaries separating 'trash talking' from cyberbullying (Ballard and Welch 2017: 438). And though in virtual environments, such as the one discussed in this paper, it is clear that, even though players rarely interact with other players they have previously played with, a general set of norms of appropriate behaviour exist.

In video gaming contexts, toxicity may incorporate behaviours such as verbal arguments over communication channels like text and voice chat, tools by which, using a microphone and a keyboard, players may interact with one another to one another (McInroy and Mishna 2017: 601). General undesirable behaviours, that virtual spaces are known for, may be classified under 'trolling', which, in Cook, Schaafsma, and Antheunis (2017: 3329), may be further categorised into *verbal,* i.e., trolling through in-game text channels, and *behavioural trolling,* i.e., the abuse of in-game mechanics with the aim of trolling. Verbal trolling may be exemplified by the deliberate taunting of other players through text and voice channels, while behavioural trolling may involve causing harm to one's own team, or the progression of the match for instance, similar to scoring an own goal or holding to the ball in football to stall time, preventing the game from progressing

further. Phenomena such as trash-talking is generally viewed by players as something ambivalent, and not outright nefarious or detrimental, and is therefore more closely associated with banter in player narratives (Ballard and Welch 2017: 436-437). Research into toxicity in online gaming have approached the topic by examining distinctly harmful behaviours such as bullying, profane language, account theft, cheating, and exclusionary playstyles that limit or hinder the chances of fully participating in the gaming activity (Teng et al. 2012). Large, multileveled, and complex entertainment software may leave numerous blind spots that go unnoticed by the game developer, enabling the exploitation of these oversights by the players, at times at the expense of others' enjoyment.

More holistic approaches towards toxicity have identified various forms of toxic behaviours, and how players direct aggression towards others within online gaming. As noted by Tan et al. (2022: 5), players may engage in emotionally or socially motivated toxic acts by hurling verbal abuse towards their own team in such games where the gameplay is organised around a cooperative setting pitting players together towards a common objective, similarly to traditional sports such as football. Toxic acts tend to, according to the authors, lower the general affect of the team and the offending player risks causing negative feelings among fellow teammates. Dichotomously users may also opt for either "passive" or "active acts", either passively choosing non-cooperation with one's own team, or actively disenfranchising one's own team by manipulating the virtual environment to disadvantage their team, by doing so jeopardising chances of victory in a MOBA for instance.

Profanity, toxic language, and anti-social in-game behaviours are rife to cause feelings of anger and frustration among players, however as research shows, the *pervasiveness,* i.e., the mere existence, of these phenomena seem to inflict feelings of anger, and not falling victim itself to profane or toxic abuse for example (Teng et al. 2012: 352-353). The prevalence of cheating or account theft is not received generally with discontent; however, the presence of profanity and exclusionary playstyles may be a source of dissatisfaction. In addition, the prevalence of toxicity seems to negatively affect the likelihood of victory, as in the case of League of Legends (LoL), a MOBA title focused on cooperative and adversarial game settings. For example, teams with

broadly toxic participant score worse on a number of in-game metrics than their controlled peers (Monge et al. 2022: 95), showing that verbal abuse and profanity may lead to lower performance overall when compared to other groups.

2.2 Factors influencing 'toxicity'

Various researchers pertain that the lack of discernible personal information, i.e., the anonymity afforded by the online platform may have a role to play in the perpetration of different toxic behaviours (McInroy and Mishna 2017: 603, Tang and Fox 2016: 514, Fox and Tang 2017: 1294). Players who tend to become targets of verbal abuse in online spaces tend to involve inexperienced or newer players who avoid retaliation (McInroy and Mishna 2017: 603). Players may also become marked for victimisation through various avenues, such as their voice, avatar, or by their gamer tag based on their perceived race by other players (Ortiz 2019: 573,577). Other research has highlighted the role of the online disinhibition effect put forward by Suler (2004). Furthermore, some authors situate the competitive gaming environments in a social climate that cultivates and rewards hostile behaviours to retain player attention, i.e., environments that provide adequate challenge and frustration (Tang and Fox 2016: 514).

Recent inquiries into toxic behaviours indicate that players who act as the perpetrators of toxic behaviour may exhibit specific motivations, as highly achievement-oriented players, those who value winning over pleasure for the sake of play, may engage in verbally abusive intrateam comments more frequently than their peers (Tan et al. 2022: 6). Moreover, research into gender dynamics and their relation to victimisation to toxic behaviour indicates that general verbal abuse may be linked to sexual harassment towards female players, though noting that the engagement, or time spent playing bares little significance in relation to prevalence of types of verbal harassment, suggesting that individuals engaging in these behaviours may be psychologically primed to act in a given manner (Tang and Fox 2016: 518).

Online spaces focused on gaming, such as Xbox Live in the study by Ortiz (2019), an online virtual gaming space and hub, have been identified by users falling under racial or sexual minority, as

spaces predominantly occupied by (in the American context), white males and teenagers, and as such presents a highly racialised and polarised space (Ortiz 2019: 578). Indeed, virtual online gaming, in the case of DotA 2 for example, may contain a bias towards males, as a number of studies examining the gaming medium lean towards a near 95 per cent rate of male informants such as in the studies by Röhlcke et al. (2018: 3), Ratan et al. (2015: 450), and Stokes et al. (2018: 334). Gaming environments in the online sphere may be conducive for fostering a system of normative beliefs regarding antisocial verbal behaviour, as suggested by Hilbert-Bruce and Neill (2020: 306), and such behaviours may be perceived as 'par for the course'. Additionally, a number of factors have been found to predict online aggression (Hilbert-Bruce and Neill 2020: 307). Player attitudes towards abusive behaviour, the age of the player, time invested and overall engagement in the game, as well as gender, male being more a stronger factor, may contribute to a heightened likelihood of engaging in damaging behaviours when gaming in an online setting, according to Hilbert-Bruce and Neill (2020: 306).

Moreover, players engaging in verbal abuse towards fellow players tend to be strangers, and not a staple of the players' regular social circle within the gaming context (McInroy and Mishna 2017: 603). Furthermore, online cyberbullying that includes verbal abuse may incorporate insults such as *pone* and *PWN,* 'to own', i.e., achieve victory over another player expressed through alternate spelling forms, and *newb,* or *noob* derived from *newbie,* i.e., someone new to the game, unfamiliar with the environment, inexperienced (McInroy and Mishna 2017: 602-603). When examining the terminology used by the offenders, one can glean that the trait ostracised of players is, as noted, the rookie status of a player, while however, the domination of another player within the game is also a source of profane jargon. Often the aggressive behaviour in the gaming environment may not even be interpreted as cyberbullying, and a mitigation of its seriousness or its effects are instead highlighted (McInroy and Mishan 2017: 603). Communities formed by players may generate their own norms, and verbal abuse, toxicity, and bullying are expected to be processed by the individual player, and players may possess both individual and community level tools to process bad behaviour (Teng et al. 2012: 352). Teng et al. (2012:352) note that for some games and genres, such as first-person shooters, a level of profanity and toxicity is socially acceptable, i.e., permitted by the player-base. According to the authors, only

two of the examined misbehaviours was positively correlated with feelings of anger: profane language and verbal toxicity, and the hoarding of in-game resources – a means of disadvantaging fellow players for one's own benefit (Teng. et al: 2012: 352-353), while cheating, account theft, and bullying are less likely to cause feelings of anger, unlike profane language or self-promotion at the cost of others' enjoyment.

As the previous studies discussed here have highlighted, toxic verbal behaviour employs jargon-like insults (McInroy and Mishna 2017: 602-603), verbal abuse is largely tolerated by the general player base out of anonymity (McInroy and Mishna 2017: 603), it is curated on the level of the gaming community at large, or players may find solutions on their own, such as in the case of Taiwanese gamers (Teng et al. 2012: 352), it is seen as a part of the culture of many gaming environments (Teng et al. 2012: 352), and the tools to perpetuate verbal behaviour may be acquired through extended engagement in the activity (McInroy and Mishna 2017: 210). In the case of DotA 2, players generally seem to have a higher tolerance for toxic and destructive behaviour, as it does not affect their engagement with the game (Lee et al. 2020: 3-4). It is worth keeping in mind, however, that while players may engage with the game, individual enjoyment of the game may not correspond linearly. In much the same manner as McInroy and Mishna's findings, the results by Mattinen and Macey (2018) indicate that a younger person may become desensitised and be less sensitive to online misdemeanour.

2.3 Multiplayer online battle arena (MOBA) and DotA 2

In this section, a groundwork for the examination of the linguistic behaviour in the corpus will be laid out. First, a basic overview of the communicative environment will be provided. Afterwards, a number of case studies will be presented to further illustrate the research conducted on the topic. The section provides a comprehensive survey on the breadth of research on MOBAs. While the focus of this present paper is on the MOBA *DotA 2*, papers dealing with the genre in a broader sense have been selected, as DotA 2 does not distinguish itself in any facet relevant to the objectives of this paper. Therefore, the following section details papers from a multi-disciplinary selection of studies ranging from sociology to linguistics.

MOBAs provide a ripe ground for interested researchers for two reasons. They bolster a massive player base and generate vast quantities of available data for a researcher to analyse (Mora-Cantallops and Sicilia 2018: 128). APIs, software allowing access and retrieval of software data (Mora-Cantallops and Sicilia 2018: 128), provide the basis for data collection in the present study as well. A literature review by Mora-Cantallops and Sicilia (2018: 130) finds that a majority of academic interest in the topic of MOBA games is focused on the *gameplay* aspect of the game, predicting match outcomes and designing modelling software. Zhang et al. (2021: 519), for example examine the assessment of "real-time win rate" based "recommended hero combination to the enemy line-ups". Studies examining the aspects of machine learning and out-come prediction are various, as pointed out earlier, however papers on more varied topics such as the civic engagement, political extremism, and the larger ecology of gaming are examined by a number of authors as well. Though scholarly interest has been expressed through various research papers, a marked lack of papers on the linguistic ecology, aside from a paper by Kwak and Blackburn (2015), in online spaces such as the one in the present paper, provides a glaring need for a linguistic analysis.

The MOBA-genre has been investigated by a number of authors. According to De Mesquita Neto and Becker (2018: 10) MOBA games, short for *multiplayer online battle arena*, have become the most popular single genre of online gaming, having ranked the most played online games in 2015. In its simplest terms the MOBA-genre can be defined as "subgenre of real-time strategy games" where two teams consisting of five players each assuming the control of a single playable character, or *hero,* attempt to destroy the enemy base (Mora-Cantallops and Sicilia 2018: 128). Kordyaka et al. (2020: 1083) define MOBAs as "subgenre of real-time strategy videogames [...] fusion of longer existing game genres such as action, role-playing, and strategy". As De Mesquita Neto and Becker point out (2018: 10), these games are highly competitive, and rely on effective interactional and cooperative skills by the players to win. Planning and strategizing, that is, and common goal-oriented action is crucial in victory (Korydaka et al. 2020: 1083). Xia et al. (2019: 501) characterises the core aspects of the game (parentheses and italics added for clarity by the author) in a succinct manner as: "a 5v5 RTS (*real-time strategy*) game in

which participants position and manoeuvre units and structures under their control to secure areas of a map and destroy the opponents' assets." Likewise, Xia et al. highlight the focus on cooperation and coordination between team members, and again stress the binary goal of the match - to defend one's own base, and to capture the other's base (Xia et al. 2019: 502). The competitive nature of the game is further emphasised by the ranked game mode, where each outcome of a match either subtracts or adds to a player's overall score, which "represents their overall level" (Kordyaka et al. 2020: 1083).

Defence of the Ancients 2 (henceforth Dota 2) is a freely distributed online game where teams of five players queue up in a match-based session (Rölcke et al. 2018: 4). The game boasts remarkable complexity in its structure, and to participate players must memorise extensive lists of in-game 'items' and 'abilities, making the time commitment required to fully participate taxing (Mattinen and Macey 2018: 3). To exploit the system to its fullest, players are required to possess in-depth knowledge of the minute interactions between the abilities of the player avatars and the various interactions between other in-game elements such as items (Röhlcke et al. 2018: 4). As a chess player, for instance, knows the names and movements of each piece, and the strategy to win, so does a player of Dota 2 intimately know the "pieces" and their interactions with one another, the rules that govern their movement, and how to emerge victorious over the opponent. In order to win, players must acquire resources, such as experience and gold, so that they may advance to and destroy the enemy base, thereby winning the match (Katona et al. 2019: 1) or destroying their "assets" (Xia: 2019: 501).

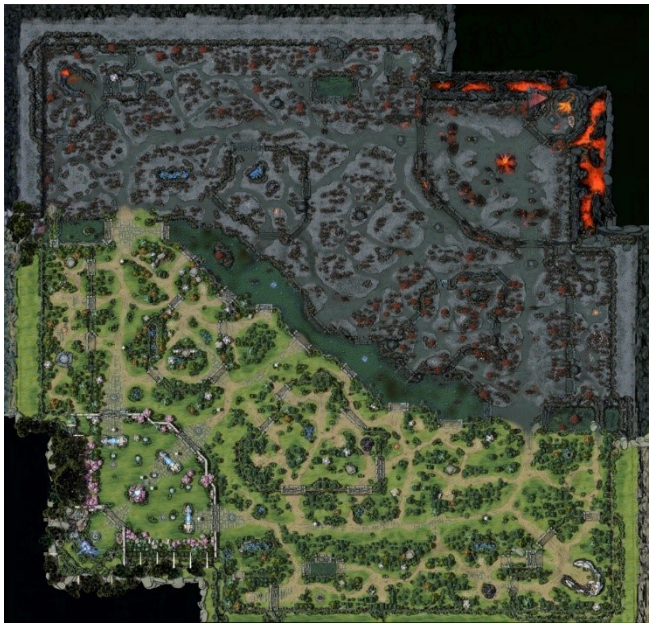

Figure 2. The 'map' of DotA 2, i.e., the proverbial chessboard[1]

Players may interact with the game by picking a hero, an in-game avatar that determines the expectations and duties of a player within a given match, such as healer or attacker (Stokes et al. 2018: 333). Each playable character possesses a unique skill set (Jarret et al. 2021: 103). Some research suggests that avatar selection itself may alter a player's verbal output for instance (Sengün et al. 2022). They noted that in League of Legends players who attack-oriented in-game roles, humanoid player characters, as well as male characters tended to increase the number of messages sent and the toxicity of the players verbal output, whereas support-oriented roles and female player characters tended to be more loosely associated with toxicity (Sengün et al. 2022: 7-8). The authors note that toxic behaviour in particular may be more strongly associated with the selected player character (Sengün et al. 2022: 8-9).

Various studies in the gaming context have underlined the core skills a team is required to possess to win. While Katona et al. (2019: 1) note that there is currently no single tactic or strategy to win, instead teams are required to coordinate their actions and predict the opposing team's intents, and subsequently adapt their strategies. Indeed, coordination, communication

[1] https://dota2.fandom.com/wiki/Map

and team play are often highlighted in the literature (Kordyaka et al. 2020: 1082, Katona et al. 2019: 1, Stokes et al. 2018: 333, Röhlcke et al. 2018: 4). Furthermore, the core skills described previously are, as noted by Stokes et al. (2018: 333), scaffolded by the structural elements of the game itself, as the two teams must advance on the virtual playing field and gain enemy territory. Röhlcke et al. (2018:4) succinctly summarise the expected skill set a player is required to possess. On the one hand, a player must possess “decision making, itemization and mechanical skills”, while on the other hand a player needs to “communicate, cooperate and coordinate”. As previously iterated, social cohesion, coordination, and team-based and prosocial action is a prerequisite to winning, and therefore it is assumed that a pro-social disposition would be inconducive to toxic language.

Players are afforded the option to choose between a casual and a more competitive ranked mode where losses and wins amount to a player's overall score (Kordyaka et al. 2020: 1083). The data within this present study is comprised of ranked matches. Kordyaka et al. (2020: 1083) highlight other features of Dota 2, which are, however, largely shared features within the genre. They note that players may use various means of communication baked into the gaming interface, such as special pings, which play a voice line from the controllable character that players on either side may see, as well as brief audio-visual signals containing symbols that signal commands such as ‘Careful! or ‘Missing!’ However, the majority of communication between teammates is facilitated by a text-chat feature (Kordyaka et al. 2020: 1083). Furthermore, as their study examines the propagation of toxicity within the gaming context, they find that these features of communication present a number of avenues for toxic verbal behaviour.

A breadth of studies has analysed the gaming medium from various perspectives. Mattinen and Macey (2018) look at the experiences and perceptions of verbal abuse by younger players in DotA 2. They examine the perceptions and experiences of verbal abuse in DotA 2 related to the age of the participants. The authors gathered the data through an online survey, netting a total of 373 responses, with a heavy skew towards males (94.2% male). The results by Mattinen and Macey (2018: 5) indicate the perception of severity of in-game harassment tends to increase with

player age, as well as the likelihood of engaging in communication abuse. Similar findings were made by Lee et al. (2020), whose paper would seem to suggest that in general, players of DotA 2 may possess a high tolerance to toxic behaviours like toxicity and abusive language. Their paper compared player experiences of identity as a DotA 2 player and their time playing the game, revealing that experiencing toxicity is not likely to significantly affect a player's sense identity as a 'Dota player', or cause them to decrease their level of engagement in the activity (Lee et al. 2020: 3-4).

Additionally, Mattinen and Macey (2018: 5) conclude that younger players incur more penalties such as timeouts and chat restrictions for inappropriate verbal behaviour. The authors seem to suggest that earlier exposure to abusive verbal behaviour may desensitise younger players to verbal abuse, which may lead to increased participation in verbal abuse as the player matures (Mattinen and Macey 2018: 6). These findings indicate that the linguistic behaviour occurring in-game may be an acquired behavioural pattern, as exposure to younger players seems to correlate with desensitization to toxic verbal behaviour. Increase in player age was found to correlate with an increased likelihood of engaging in toxic communication in-game (2018: 5). Interestingly, this finding is supported by Murnion et al. (2018: 210), who examined another popular online team-based game, *the World of Tanks*, whose findings indicate that toxic verbal output is first produced by more experienced players and reproduced by younger or more inexperienced players, who become accustomed to it. They propose that cyberbullying, and by extension, toxicity, is a type of behaviour that new players pick up through participation (Murnion et al. 2018: 210). Furthermore, the study reveals that penalties such as dying in-game were associated with an increased propensity to cyberbullying remarks by players (Murnion et al. 2018: 210).

Others like Kordyaka et al. (2020) investigate Dota 2 and the broader MOBA-genre through the framework of toxicity, while authors such as Röhlcke et al. (2018: 2) examine cognitive aspects of playing the game, e.g., the relationship between working memory capacity, grit and time played affect a players skill rating. Adding to the previously referred resilience towards toxicity, the results by Röhlcke et al. (2018: 5) indicate that time spent playing may be the strongest predictor

of skill in Dota 2, whereas working memory capacity did not produce statistical effects in their study. Though, the authors note that a selection bias may be explain the results, i.e., players with high working memory capacity were picked for the study, thus skewing the effects (Röhlcke et al. 2018: 8).

2.3.1 2.3.1 MOBAs through a socio-political lens

A newly-released literature review on the prevalence of right-wing political extremism raises concerns and highlights a further need to describe the linguistics of toxicity and inflammatory language within the medium. The literature review underlines several characteristics that permeate the range of academic concern regarding the industry. The authors stress a general lack of diversity within game studios, as a majority of developers are white and heterosexual, leaving women and people of colour (PoC) bereft of major representation (Wells et al. 2023: 6). In relation to discussions of political extremist thought, Boluk and LeMieux cast light on the regional divide between western and eastern communities of Dota 2 players. According to the authors, there exist certain narratives that are perpetrated within discursive landscape of Dota 2's Esport-scene that closely define the Asian and European players based on stereotyping and caricaturing (Boluk and LeMieux 2017: 213). They draw parallels between macro-level international economic trends and strategies, and the conceptualization of the different player bases noting that "terms like "farming" and "ricing" have become common ways to describe long-term, economy-driven strategies in contrast to the micromanagement of team fighting" (Boluk and LeMileux 2017:213). Within the context of the quote, the terms 'ricing' and 'farming' are well established domain specific jargon terms. The authors bring attention to how narratives are constructed within the casting of E-Sports matches, wherein Chinese teams are portrayed as collectivist and calculated while European players are praised for "feats of daring [...] heroic individualism".

Though while such narratives may permeate the professional scene of Dota 2, Ismangil (2019) casts light on nationalist rhetoric within Chinese online communities constructed around Dota 2. Ismangil's findings on memes in Chinese Dota 2 communities corroborate findings by Boluk and

MiLeux by demonstrating that larger macro level cultural and social attitudes may be reflected on the level of communities of practice. Ismangil draws parallels between Chinese state policy, the promotion of Chinese unity through nationalist policies, and the engagement of the Chinese Dota 2 community in the Esport scene, arguing that propagation of nationalist memes within these communities may reinforce tenets set by the Communist Party (CP). Memes such as *diaosi* (Ismangil 2019: 238) act as veils through which state policies and disgruntlement with the larger system is expressed, while memes such as *cai, liu, shen* or 3154 provide insight into the self-referentiality and intertextuality within the E-sports following public (Ismangil 2019: 239). Ismangil argues that through these memes Chinese nationalist attitudes are reinforced by framing the viewer to engage with the ideas conveyed through the imagery (Ismangil 2019: 241) Such nationalist tones may be additionally reinforced by regionalist wishes of dominating international tournaments (Sweeney et al. 2021: 69), as both Dota 2 and League of Legends are played on a competitive scene and boast annual championships like *the International* (henceforth *TI*) and the League of Legends Championship series (henceforth *LCS*) – a notion highlighted by Ismangil as well.

Stokes et al. (2018) examine the relationship between playing MOBA-games and engaging in civic actions, such as protests. The authors suggest that increased engagement in micro-social environments and whether a pro-social disposition within the gaming context may be more likely to engage in civic action (Stokes et al. 2018: 328). They note that players opting for cooperative choices within the game were 3.7 per cent more likely to take part in a protest than their less cooperative peers. Furthermore, an investigation of hours spent playing reveals that between moderate, high and extreme hours spent gaming, only extreme hours spent playing seems to affect the likelihood of protesting (Stokes et al. 2018: 339).

While the majority of the papers cited here involves a heavy male-leaning bias in their data, female perspectives on gaming have received fewer papers. Within the relevant research on the MOBA-genre Ratan et al. (2015) examine female perspectives on *League of Legends* through two studies. Study 1 revealed that both male and female players are intimately familiar with the volatile and toxic character of the gaming environment, with females reporting being

pigeonholed into the support role (Ratan et al. 2015: 443-444). As the moniker suggests, players in the support role are heavily dependent on other members of team, and instead of focusing on attacking the enemy, they support their own team, and are, as one might term it, more reliant to the more 'meaningful' classes. In their second study the authors report male participants having a generally higher match count and a higher average skill level, based on the ELO-ranking used to track a player's skill level in relation to other players. However, the authors note that gender does not seem to predict skill level. Furthermore, female players were more likely to play the support role, which was predicted by co-play with a romantic partner (Ratan et al. 2015.: 452-456). The findings highlight the stark male bias of the gaming environment, and point toward a rigid hierarchical structure, which players are expected to follow.

Others have examined the gaming context from a unique perspective. Jarret et al. (2021) examine the MOBA League of Legends from the perspective of affective economies, while authors such as Sweeney et al. (2021) investigate Esports through gambling markets formed around the tradable virtual goods provided by the games' publishers. Jarret et al. (2021) examine the relationship between the developer of League of Legends, Riot Games, and the consumer-base of the MOBA. While the papers are only loosely connected to the present paper, findings such as Sweeney et al.'s are worth reporting. They report that certain biases favouring European players interpreted as underdog performers may be over-betted on a statistically significant level (2021 83).

2.3.2 Linguistics of the MOBA-genre and online gaming more broadly

Fewer papers commenting on the linguistic devices used to perpetrate toxic language have been produced. In Kwak and Blackburn (2015) the authors examine a million-word corpus on toxic chat messages produced in LoL. It is worth noting that the corpus used in the study is primarily focused on the toxic messages sent by players and does not represent the whole linguistic landscape of the medium. Their analysis reveals that 'toxic players' tend to produce longer messages compared to their 'non-toxic' peers (3.139 words to 2.732 words) (Kwak and Blackburn

2015: 212). Comparing uni- and bigrams (constructions consisting of one or two items) produced by toxic players and typical players, it becomes apparent that general insults and expletives are highly typical of toxic players' linguistic behaviour, see table 1. The authors comment on the frequency of the unigrams, observing that 'fucking' itself is not a unique item, whereas many of its orthographic variations are (Kwak and Blackburn 2015: 213).

Table 1. Ten most frequent uni- and bigrams in Kwak and Blackburn (2015: 213).

Frq.	Unigrams	Bigrams
1	retards	fucking retard
2	nigger	report noob
3	garbage	fking noob
4	uninstall	fucking useless
5	piece	fuck team
6	pathetic	report fucking
7	fuckign	stupid noob
8	fukin	pussy ass
9	nooob	play bots
10	bots	play fucking

A comparison of player groups also reveals that the set of unigrams produced by typical and toxic players vary in size, with toxic players producing a total 80, while typical players produce 91 unigrams, indicating a more limited vocabulary (Kwak and Blackburn 2015: 215). Furthermore, the study illustrates other facets of toxic player communication: toxic players in League of Legends rarely use emoticons or affect signifiers, nor do they use apologies. The use of coordinative in-game 'call outs' and movement coordination related communication diminishes over the duration of a match. Finally, the authors note that toxic players stop using appraisals like *gj* (good job), after a point in the game, similarly, to coordinating messages (Kwak and Blackburn 2015: 216).

2.4 (Im)politeness, face, and face-threatening acts (FTAs)

As the present paper relates to impoliteness research, this section will provide an inspection into the literature of linguistic pragmatics, detailing the theory of politeness as described by Brown and Levinson (1987), as well as provide the groundwork for the analysis of face-threats in the present paper. Further connections will be established between Brown and Levinson's work by examining influential (to whom) concepts in politeness research, such as Rapport management by Spencer-Oatey (2002).

2.4.1 Politeness and impoliteness

The way politeness is conceptualised in academic research differs from lay understandings in a number of ways. Based on the distinctions of positive and negative face, Brown and Levinson (1987: 317) derive a framework of politeness following the same distinction. So called *positive politeness* concerns with respecting the interlocutor's positive face needs, i.e., the need to have one's face needs validated by other members of the community, or the co-locutor. Positive politeness requires that the speaker indicates to the hearer that their wants are desirable to the speaker as well and are thereby validated. It follows that face attacks should be minimized by assuring the hearer that their wants are shared by the co-locutor, or that the hearer to one degree or another, is important to the speaker, and by extension, the perception of his or her desirability is reinforced. Positive politeness serves to assure the hearer that their positive face is not brought into question during a communication event. Leech (2014: 11) further develops this dichotomy of positive and negative politeness into pos- and neg-politeness. In Leech, positive politeness, or pos-politeness, is conceptualised as actions or utterances that "assign positive value to the addressee" (Leech 2014: 12). Leech further characterises pos-politeness strategies such as apologies, expressions of gratitude as corrective manoeuvres speakers may undertake to remedy social imbalances, such as debts of gratitude (Leech 2014: 12).

While positive politeness concerns with the preservation and attendance to the hearer's positive face needs, or rectifying socio-pragmatic imbalances, negative politeness strategies aim to preserve the hearer's or co-locutor's negative face. Negative politeness involves knowledge and acknowledgment of the hearer's "territory" and the maintenance of internal autonomy in the hearer, their ability to act without imposition from outside actors (Brown and Levinson 1987: 317). Negative politeness is, therefore, primarily concerned with leaving the hearer the ability to retreat from any form of imposition (Brown and Levinson 1987: 317) like requests, where the hearer is expected to either reject or accept the duty predicated by the request in question – avoiding an FTA (a *face-threatening-act*). In Brown and Levinson linguistic negative politeness is characterised by a level of "formality and restraint", which may be realised through the use of the passive voice, use of hedging devices, and other softening mechanisms. For example, in a restaurant setting, a customer may be encouraged to leave the service staff a tip, an expression which might be worded as: ***It is usually*** *customary to leave a tip.* A more direct approach would be to employ a request such as: *Please leave a tip.* Leech (2014: 11) highlights a further distinction between negative politeness and neg-politeness; neg-politeness strategies aim to, instead of protecting the hearer's "territory", remove and mitigate the factors that might offend the interlocutor. As Leech notes, common ways to perform neg-politeness involve strategies such as indirectness, use of hedging devices, similar to negative politeness by Brown and Levinson.

Impoliteness, as a separate concept from *politeness,* may be conceptualised as negatively received ways of behaving in particular situations. Beliefs about the impoliteness of an act are determined by the incongruence between what is expected and appropriate behaviour in a situation, and the act itself. Acts and behaviours that go against the 'script' are expected to cause emotional fall out and are thus interpreted as to cause offence to the hearer (Culpeper 2011: 23). According to Culpeper (2011: 22), impoliteness is comprised of held beliefs regarding appropriateness of certain behaviours in particular situations, and the attitudes and evaluations that follow the deviation from these expectations. Diverging from pragmatic analyses, impoliteness may be defined through the presence, and conversely, absence of politeness markers (Kerbrat-Orecchioni 2013: 20-21). Furthermore, the unlikely absence of these politeness markers facilitates 'negative impoliteness' such in the omission of conventional politeness

markers, e.g., greetings, apologies, or verbal expressions of gratitude. Conversely, incorporating markers of impoliteness comprises 'positive impoliteness (Kerbrat-Orecchioni 2013: 20-21). Though an action or utterance may be made impolite by deviating from context dependent norms, a number of linguistic formulations to deploy impoliteness have been proposed by authors.

Aside from purely pragmatic standpoints, according to Culpeper (2013: 8) linguistic impoliteness may be realised by *vocatives* (e.g. 'idiot', imbecile' etc.)*, personal negative evaluations* (e.g., 'you idiot, imbecile, etc.)*, silencers* (e.g., 'shut the fuck up')*, threats,* or *personalised third-person negative references* (e.g., 'she/he is an idiot') (Culpeper 2011: 135). Culpeper notes that, although speakers possess overt means of performing impolite acts, speakers tend to prefer implicit or covert means of performing impoliteness (Culpeper 2013: 9). Impoliteness, according to Culpeper, may integrate a "playful frame", and as he notes, impoliteness possesses a certain level of creativity (Culpeper 2013: 9). As impoliteness is highly context dependent (Culpeper 201: 9), and has the possibility to incorporate elements of playfulness, it follows that to discern between "mock-impoliteness" and 'malicious' impoliteness, a keen eye is required. Moreover, the notion of impoliteness, or the lack of politeness, can be extended, as Kerbat-Orrechioni distinguishes several notions linked to impoliteness in the form of *over-politeness, non-politeness,* and *polirudeness* (Kerbat-Orrechioni 2013: 20-21)*.* Over-politeness is conceived as the overuse of politeness indicators with respect to the norms governing the interaction, non-politeness is then, the lack of explicit markers, whereas polirudeness denotes face-threats couched in polite discourse, a superficially polite utterance containing a face-threat.

2.4.2 Face and face-threatening acts (FTAs)

Early research defines face as the "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact"(Goffman 1967: 5), encapsulating the idea that an individual's assessment of the *self* interacts and affects the perceptions and judgements made by others in a given situation (Culpeper 2011: 25). In Brown

and Levinson, *face* is defined as the basic set of wants of every individual in a community, that every member knows other members desire or perceive as desirable (Brown and Levinson (1987: 312). They divide face into two subcategories that define various aspects of face; namely, the negative and positive face that speakers hold (Brown and Levinson 1987: 312). Positive face describes the wants of each individual to be desirable to other members of a given community, whereas negative face is concerned with the individual desire to have one's autonomy not be challenged by other members of the social group (Brown and Levinson 1987: 312). The notion of face extends the idea of positive and negative politeness into the practical discourse situations and operationalises the concept to enable more holistic analyses.

2.5 Other approaches to *face* and *face-threatening acts*

Spencer-Oatey develops the idea of rapport management as consisting of 'face and sociality rights' (Spencer-Oatey 2002: 540). Her characterization of face borrows heavily from the work by Goffman (1972), and bears striking similarities to the theory of politeness by Brown and Levinson in the dichotomous interpretation of 'face', as the "value a person effectively claims for himself by the line others assume he has taken during a particular contact", i.e., "face" is the value by which a speaker conducts themselves in a given interaction. *Sociality rights* are defined as the "entitlements" that an individual expects in his or her interaction with others (Spencer-Oatey 2002: 540), what might be termed as 'equal treatment' between peers. As Spencer-Oatey highlights, *face* relates to the perceived value that an individual holds for themselves in relation to others, while sociality rights are the notions of fairness and equal treatment with their peers (Spencer-Oatey 2002: 540).

Spencer-Oatey categorises the concept of *face* into four categories: *quality face*, *social face*, *equity rights,* and *association rights* (Spencer-Oatey 2002: 540-541). Quality face, in accordance with previous characterisation by Brown and Levinson for example, holds that individuals fundamentally desire to be seen in a positive manner, and to be acknowledged for the admirable traits or qualities they may possess. Social identity face governs the desire to be recognised in a

given social role or identity, e.g., professional or personal, i.e., the public worth, that an individual may claim. Whereas Brown and Levinson (1987: 317) interpreted face as consisting of negative and positive face, Spencer-Oatey does away with the concept of negative face in favour of *equity rights* as the freedom from imposition, though both concepts approach the topic from a similar viewpoint. She expands this conceptualization through *cost-benefit* and *autonomy-imposition axis*, where the former signifies the ratio of exploitation experienced by individuals to the need for reciprocal exchange of politeness, i.e., how far individuals are willing to withhold the reciprocal bargain on their end. Autonomy-imposition axis relates to the "extent to which people control us or impose on us" (Spencer Oatey 2002: 540). It describes the degree of personal autonomy and outward pressure placed upon us. According to the author, relational management, the management of intersocial relationships, is then comprised of four main aspects: face and rights, autonomy and cost-benefit, association and autonomy axis, along with miscellaneous interpersonal, inter- and intragroup orientations (Spencer-Oatey 2002: 539).

When operationalised into the gaming environment particularly the types of face outlined by Spencer-Oatey may be characterised through ability excel in the game and triumph over others, i.e., as the 'good player' - the quality face. Similarly, we may characterise the social face players maintain as 'valuable members' of the team, 'contributing members' of the team, or based on the in-game roles and how well each player fare in their selected role. Within the virtual environment equity rights or the freedom from imposition may be conceptualised as the expected trust and capability that is afforded to each player, i.e., that one is capable of filling their role, whereas cost-benefit analysis by players seems less relevant. As Lee et al. (2021: 3-4) note, players of Dota possess a high tolerance to abusive behaviours, and may therefore signify a high tolerance to exploitation at the cost of their own need for polite reciprocity.

3 Cursing, swearing, and 'taboo language'

This section provides a description of a number of studies that have previously examined swearing. Additionally, many of the studies comment on the topics of the present study, i.e., the

lemma *fuck,* such as the study by McEnery and Xiao (2004). Other research papers have looked at instances of swearing in Irish English (Schweinberger 2018), while papers such as Martinez and Petrejo (2012) have delved more deeply into the use of taboo words by groups such as teenagers. This section will first establish proper terminology for the paper, and subsequently survey in detail a number of studies describing the distribution of profanity along speaker groups, as well as some of lexical items used by speakers.

3.1 Establishing terminology and defining concepts

Swearing as defined by (Love 2021: 742) refers to language use pertaining to socially opprobrious taboo topics such as religion, bodily functions, or procreation that can be employed situationally to express a variety of social functions such as humour, emotion in exclamations after physical injury, e.g., *Shit!*, or verbal abuse, where the intent is emphasised through expletives and vocatives, e.g., *You damn moron!*. Though others like Jay (2009: 154) maintain that curse words and swears in English may be categorised sexual, religious or blasphemous, scatological and excretory functions, or norm-breaching 'deviant' behaviours or features such as disabilities, slurs and other remarks that include race, gender, or sexuality, as well as colloquial and vernacular vocabulary considered taboo. Likewise, other such as Sulpizio et al. (2019: 84) employ the term 'socially opprobrious words', which in a similar manner highlights the social stigma associated with the use of a particular set of lexemes.

Others such as Coats (2021) employ a general descriptor of 'bad language' to describe language use colloquially referred to as swearing. Similarly to Love and Sulpizio et al., Jay conceptualises curse words and swear words as the "lexicon of offensive emotional language" (Jay 2009: 153). Jay and Janschewitz (2008: 269-270) introduce a distinction into the swearing within their study. They highlight the differences between propositional and non-propositional swearing, noting that propositional swearing denotes deliberated swearing in the service of a social function such as building camaraderie. Moreover, the speaker is placed in control of the utterance and to achieve a premeditated result, whereas swearing in the non-propositional

context underlines a lack of control over the production of the offending utterance, stemming from spontaneous emotional outbursts. Propositional swearing in Jay and Janschewitz is underlined by a socio-pragmatic function, while non-propositional swearing highlights non-volitional, spontaneous swearing in response to emotional stimulus. When asked to generate taboo words, a majority of speakers tend to produce words associated within the categories of sex, excretory functions, and demeaning names, i.e., bastard (Jay and Jay 2015: 257). As the taboo language itself employs a general sense of norm breaching to achieve socially or emotionally motivated goals, and involves a lexicon used to describe fringe elements or stigmatised human behaviours that are viewed as unfavourable or as something that ought be avoided, the present paper shall defer to the term 'taboo language' to describe the object of study, as thus far.

Jay and Jay (2015) provide counter evidence to the folk belief that high swearing individuals may possess a sparser vocabulary than their less frequently swearing peers. It was found that in a laboratory setting, speakers were able to generate more words classified into categories such as animals than they were able to generate swearwords, taboo words, and slurs (Jay and Jay 2015: 254). Speakers were able to produce more lexemes outside of the taboo word category both orally and in writing (Jay and Jay 2015: 255). Furthermore, the taboo words produced by speakers tend to mostly reflect a limited set of lexemes, as the ten most frequently encountered taboo words comprised 60 per cent in oral production, and 50 per cent in written production, indicating that speakers may favour a seemingly set selection of taboo words (Jay and Jay 2015: 254-255). Moreover, speakers tend to avoid slurs (Jay and Jay 2015 254, 255). They find that some of the categories from which speakers harness their taboo language consist of scatological references, sexual references, and 'pejorative names' such as *slut* (Jay and Janschewitz 2008: 257). Speakers generally employ fewer racial and gendered slurs (Jay and Janschewitz 2008: 257) a tendency which is echoed on CMC interaction on Twitter (Coats 2021: 50) - notions echoed by previous studies, as well.

A study on contextual variables affecting swearing likelihood, and perceived offensiveness of the activity, provides evidence indicating that a speaker's perception of the heinousness of a word is

significantly associated with the subjective speaker, where the offending utterance was produced, and how taboo the word is received, i.e., the normative evaluation by recipient. Additionally, the authors find that statistically significant relationships between speaker and location, location and tabooness, as well as tabooness and speaker further describe the phenomenon (Jay and Janschewitz 2008: 278-280). Furthermore, they note that factors affecting the likelihood of swearing of speaker, location and tabooness were significant determinants, as well as their associated relationships speaker and location, location and tabooness, and tabooness and speaker (Jay and Janschewitz 2008: 278-280.) There is reason to believe that interlocutor presence may determine some of the offensiveness behind an utterance, as an utterance addressed indirectly may be intended to be more offensive than a direct one, i.e., whether the addressee is present (Stenström 2017: 174). In addition, comparisons between native speakers and L2 speakers indicate that, while no direct relationships can be observed, native speakers tend to provide more nuanced accounts of taboo word offensiveness. Later acquisition of fluency of English among L2 speakers seems to affect the offensiveness ratings of taboo words, with late L2 speakers of English rating taboo words more offensive than their early counterparts (Jay and Janschewitz 2008: 283).

3.2 Swearing as English L1

An examination into the diachronic changes within British English from the 1990s until the 2010s using the BNC corpus reveals that a general decline in the popularity of swearing has taken place (Love 2021: 749). The overall portion of swearwords to other words is quite low in both time periods, however, only making up 0.23 per cent in the 1990s, and 0.14 per cent of tokens examined (Love 2021: 749). Three swearwords, in particular, emerge as the core of BrE taboo language, namely those of *fuck*, *bloody*, and *shit* (Love 2021: 750). When comparing the distributions of these items across the 1990s and into the 2010s, a distribution emphasising a growing popularity of the latter, with the first retaining its level of usage across the examined periods, while the use of *bloody* among speakers of BrE has seen a decline since the 1990s (Love

2021: 750). As a matter of fact, Love notes that *fuck* has become the more dominant of the three variants (Love 2021: 751).

Further inquiry into the linguistic profile of BrE swearword and taboo language use reveals yet more numerous examples when contrasted with AmE. Generally, BrE speakers tend to deploy roughly an equal measure of swearwords with speakers of AmE, with higher frequencies of swearing among colleagues suggesting that swearing in BrE is more socially acceptable in workplace environments than in AmE, even among those with a higher education background, reflecting cultural differences (Dewaele 2015: 323-324). There is additionally a noticeable gap in the perception of offensiveness between certain 'emotinally-laden' words among speakers of the two varieties, as out of the studied lexicon American speakers provided higher offensiveness ratings for more words than did the British speakers (Dewaele 2015: 323-324). Though the American speakers seem perceive the offensiveness factor slightly more severe than their British counterparts, the American speakers employ the lexicon more frequently by comparison (Dewaele 2015: 330).

3.3 Cross-linguistic analyses and developments in swearing research

Studies by McEnery and Xiao (2004) and Martinez and Petrejo (2012) examine the lemma *fuck*. The findings of these studies indicate that *fuck* is much more frequently observed in spoken language, with pronounced frequency in the case of *fucking,* the present participle (McEnery and Xiao 2004: 238). Additionally, the frequency of the word was higher in younger cohorts in both spoken and written data by ages 15-24 and 24-35 (McEnery and Xiao 2004: 242, 250). A similar finding was produced in Martinez and Petrejo's examination into the COLT and SCoSE corpora, where it was confirmed that teenagers' use of *fucking* and *bloody* where almost nine times more likely in teenagers when compared to adults in casual spoken corpus data (Martinez and Petrejo 2012: 782).

Cross-linguistic comparisons of teenage swearing between English and Spanish reveal that both groups of speakers possess nearly identical lexical inventories, semantically. Spanish teenagers employ a relatively greater number of swearwords than their English peers, with three most frequent lexemes in each speaker group being *fuck* or *fucking*, *god*, and *shit* for English teenagers, and the Spanish equivalents *joder/jo/jodido*, and *puto/a,* with *coño* or the English *cunt* for Spanish teenagers (Stenström 2017: 160). It is worth noting that the Spanish *mierda*, the English equivalent for *shit* is still ranked fourth in the ranking (Stenström 2017: 160). Stenström notes, however, that many of the semantic features and functions thought to permeate swearword usage may have undergone processes of semantic bleaching and pragmatic strengthening, losing some of the tabooness generally associated with them, as well as their former associative meanings, such as the case of *god* where the religious meaning has become more transient (Stenström 2017: 175). For instance, the Spanish use of *hijo/a de puta* and the English *god*, are generally not associated with their literal or prototypical meanings, as *hijo/a de puta* seems to, on the one hand refer to a disliked set of features, and on the other hand, serve as a form of rapport building, as an intimate form of speech among peers (Stenström 2017 168), and as such highlights the social function of swearing in itself.

Looking at examples of individual swearwords as lexemes, *fuck* has been thoroughly incorporated into various languages, with earliest written attestation in Norway dating back to 1948, with Denmark and Iceland adopting the word during the late 1960s and early 1970s, with Russian adopting the word form in the late 1970s (Vatvedt 2019: 101). The Germanic languages seem to have adapted and assimilated the English lexeme both grammatically as well as having developed orthographic approximations suited to suit the target language phonology (Vatvedt 2019: 101). Conversely, while the Germanic languages seem to have adapted the word more holistically, the Russian loanword variant has been assigned to a new grammatical role, acting as a noun instead and acquiring declension patterns of a noun, i.e., Russian cases (Vatvedt 2019: 101). Furthermore, semantically, the adapted forms seem to serve similar roles to their English counterparts.

All three Nordic languages have adapted the English semantic notion behind expressions such as *fuck you*, designating rejection or admonition, largely resembling the phonological countour of the English variant (Vatvedt 2019 101). Additionally, while Icelandic and Danish have adapted the exclamatory form *fuck* and *fokk*, respectively, the semantic profile of the word in Norwegian seems to incorporate more idiomatic expression in *føkk opp* with the meaning 'to mess up, to flounder' (Vatvedt 2019: 101). The Russian variant 'fak c nim' is beholden to a more limited semantic range than its Nordic counterparts by confining its range in the English 'fuck it', serving as a negatively loaded adverb of stance, and would seem to convey a broader semantic meaning of 'dismissal'. Regards tabooness, the Nordic speakers deem the various forms of the word from 'relatively taboo' to 'context dependent', similarly to Russian speakers (Vatvedt 2019: 101. Stylistically, the forms are deemed highly informal in both Russian and Icelandic, whereas linguistic norms in Norway and Denmark may allow the use of the word and its variants to be employed in written communication, and in Denmark even in children's TV broadcasting (Vatvedt 2019: 101.).

In a study by Rosenberg et al. (2016: 309) the authors identified the most frequently used curse words used by the participants in their set of two studies on cursing and affect. By comparing the results of their study to Google n-grams, the authors provide a list of the most frequently reported curse words by the (American) participants: *cunt, fuck, motherfucker, asshole, pussy, fucker, nigga, faggot, slut*. The study finds that speakers favour a set of established curse words, as the 30 most frequently reported words accounted for three thirds of the words produced in the study. The authors note that this list only contains singular words, possibly due to questionnaire design (Rosenberg et al. 2016: 310). Contrasting between gender of the participants reveals that females rate taboo or curse words as more taboo than their male peers, though females and males produce curse words at similar frequencies (Rosenberg et al. 2016: 312).

Though, females and males may prefer different sets of cursing lexicon, as in the Nordics, for instance, females tend to employ profanities directly stigmatising certain female behaviours and characteristics, types such as *bitch*, *boob* and *slut* may be overused by female speakers (Coats

2021: 47) Additionally, it was found that the more 'intense' a curse word was rated, the lower its frequency of use was reported (Rosenberg et al. 2016: 312). When slurs and other more potent profanities, such as references to homosexuality, are used, it is more likely that a male speaker may be the producer of the utterance, though slurs such as ethnic and racial terms are relatively infrequent on sites like Twitter for instance (Coats 2021: 50).

3.4 Effects of age, gender, and education

A number of studies attest to the differentiated gender distributions within the dissemination and deployment of swear words. Vatvedt et al. (2019: 101) note that the usage of loanwords of the English *fuck* in Norwegian are mostly propagated by younger male speakers of lower social class, by younger and middle-aged speakers in Danish, by teenagers and young adults in Iceland, and, interestingly, by younger female speakers amongst Russian speakers. Schweinberger's (2018) findings regarding the use of curse words in Irish English reveals similar findings, indicating that in Irish English 26–33-year-olds use more curse words than the younger and older cohorts (Schweinberger 2018: 11). Though his study also found that 19–25-year-olds used the second most curse words. Likewise Love (2021: 752) attests to the findings by McEnery and Xiao, as he notes that there is a steady decrease in the use of swearwords after a speaker's twenties, prior to which the frequency of use in childhood is increased. This is also corroborated by Schweinberger's findings. Both studies by McEnery and Xiao and Schweinberger confirm the finding that male speakers may produce curse words and taboo language more frequently (McEnery and Xiao 2004: 240, 248, Schweinberger 2018: 11), as corroborated by findings from Stenström (2017).

De Klerk (1992: 283, 286) provides contrary evidence of the gender dynamics active in profane language production among youths, as her findings suggest that not only do teenage female speakers produce comparable numbers of swearwords as their male peers, but the lexical range of profanities of these speakers (males and females) are quite similar to one another. Her findings also indicate that sex-specific profanities and taboo language are diverse, as not one

gender was emphasised in the production of swearwords by the participants (De Klerk 1992: 288).

Conversely, in the study on Irish speakers of English, male speakers produced taboo language nearly 2.7 times more frequently than their female peers (Schweinberger 2018: 11), and *fucking* appearing 1394 in male speech when compared to the frequency among female speakers (321 instances) (McEnery and Xiao 2004: 240). Similarly in the study by Love (2021: 751-752) it was found that the use of swearwords is nearly twice as frequent among males when compared to females in the 1990s, on top of which female speakers were reported as using fewer "pure swearwords". Within the trend of decreasing swearword usage in the 2010s males firmly retain the frequent user status, with 18.15 words per thousand for male speakers and 10.81 words per thousand for females (Love 2021: 751-752). Additionally, controlling for the gender of the recipient, or interlocutor, seems to produce different results yet. In McEnery and Xiao (2006), instances of the lemma *fuck* were more frequent with intended male audiences and same-gender audiences as mixed-gender conversations included fewer overall instances of *fuck*. When conducting cross-linguistic examinations, some studies suggest that males tend to more frequently make use of swearwords and produce more numerous instances of swearwords overall, though specific lexemes may be overused by female speakers, such as the words *bloody* and *god* (Stenström 2017: 161-162). For example, both Spanish and English teens produce more overall swearwords when the speaker is male (Stenström 2017: 161-162).

Gendered examinations also reveal that female speakers may perceive taboo words such as swearwords as more offensive than their male peers (Jay and Janschewitz 2008: 283) which may encourage female speakers to refrain from using taboo words of higher perceived offensiveness. Moreover, gender-based differences in swearing output may also differ when comparing L1 and L2 linguistic output, as speakers tend to overuse L2 swearwords in English, with male speakers producing an increased number of swearwords in English (Coats 2021: 39-40). In Nordic countries, as observed by Coats, discrepancies between swearword distributions are most starkly exposed in Finland, and the least pronounced in Sweden (Coats 2021: 38-39).

As established, gender-based differences in swearing frequencies differ greatly by region. Coats (2021) provides a further breakdown of Nordic swearing. According to Coats, Icelandic male speakers tend to favour religious profanities referring to Hell and God, while words designating stigmatised female behaviours and characteristics are favoured by female speakers, though the English equivalent *cunt* is still proliferated by male speakers. Norwegian profanities are dominated by male speakers, employing a traditional lexicon of swearing, i.e., "Devil- and Hell-related words" (Coats 2021: 46). Conversely, Danish swearing is characterised by overuse by female speakers, producing most frequent word types, whereas Danish males tend to favour the intensifiers *sgu* and sgi (Coats 2021: 46). Similarly, to Norway, Swedish males tend to favour the traditional swearing lexicon with references to heaven, hell, devil etc., while Swedish female speakers, excluding the intensifier *jävla*, show a preference to euphemisms "mild swears" in place of 'hard swears' (Coats 2021: 46-47). Again, lexicon stigmatising female behaviour is favoured by females, i.e., *whore* or *cunt*. Finnish speakers exhibit the largest disparity between male and female speakers in terms of the frequency of use of swearwords, with males showing marked overuse with words such *cunt* and *ass*. Finnish females then favour female-indicating terms such *chick* (Coats 2021: 46-47). Overall, male speakers tend to employ the 'traditional' native L1 lexicon consisting of references to religious concepts such as God and Devil, as well as references to bodily functions, genitalia, and sexual acts in the L2 English (Coats 2021: 47). In comparison to males, females employ more euphemisms, and show a clear overuse in lexicon referring shunned or stigmatised female coded sexual behaviour, such as promiscuity or homosexuality (Coats 2021: 49).

Furthermore, in McEnery and Xiao, the education level of the speaker seems to affect the number of instances of the lemma forms of fuck. Speakers who did not pursue further education after the ages of 15-16 produced the highest frequencies, decreasing as the level of education rises – indicating an inverse correlation (McEnery and Xiao 2004: 246), while Schweinberger's data does not appear to boast such an effect, indicating that in Irish speakers of English, swearing and cursing permeate the discourse more holistically. The results of the study by Love provide contradictory findings, indicating that a general drop in swearing frequencies among lower classes has decreased, though he notes that middle class speakers seem to exhibit

higher rates of swearing than other groups (Love 2021: 755). In Stenström (2017: 163) it was found that among teenagers, a lower-class background speaker in L1 English speakers may generally produce more swearwords than other groups, though the author notes that upper-class speakers are very closely behind lower-class speakers when comparing frequencies of swearing between groups. In Spanish speakers, middle-class speakers dominate over other groups when comparing swearing frequencies.

Though evidence by McEnery and Xiao indicates that speakers with lower education levels may use more curse words, it is necessary to mention the findings by Dewaele (2017), who examine the relationships between swearing and a number of sociolinguistic variables, and personality traits like extraversion in L1 and LX speakers of English. Findings by Dewaele indicate that more highly educated individuals may produce more curse words with friends and family, and the least with strangers or colleagues (Dewaele 2017: 337-338). Furthermore, L1 speakers were reported using more curse words than their L2 peers, which may result from the larger linguistic toolbox, i.e., L1 speakers have more linguistic resources than L2 speakers (Dewaele 2017: 341). Findings by Baruch et al. (2016) provide more nuance to swearing by highly educated speakers of English. Speakers in socially prestigious occupations like business executives, doctors and law practitioners seem to be highly aware of the situational appropriateness of swearing.

Professionals in leadership positions will avoid swearing to set an example and medical professionals will not swear in the presence of a patient (Baruch et al 2016: 153-154). Respondents stated that so called ‘soft’ and ‘hard swearing are present in the workplace, i.e., indirect insults and conventional swearing, and when conversing within the workplace, swearing was only ever used in face-to-face meetings, and generally not used in written correspondence (Baruch et al 2016: 154). Swearing was found to have a number of functions as stress relief, a way to express negative emotions, a source relief in a stressful environment, creating distinction between home and work life, cultivating grit, a tool to forge more informal relationships with co-workers, and from a managerial perspective, motivating employees to work more efficiently (Baruch et al. 2016: 155-156).

Based on the brief literature review provide here, a number of conclusions can be drawn from the way speakers of English employ curse words in their written and spoken productions. In McEnery and Xiao (2004) it was revealed that the lemma *fuck*, which the present paper concerns itself, is distinctly a trait of male speech, which was also confirmed by Schweinberg's results. Swearing seems to also be a feature of speech favoured by younger speakers as the findings by McEnery and Xiao, as well as Schweinberger, and Martinez and Petrejo's results unanimously corroborate. Level of education was found to be significant in at least one study by McEnery and Xiao, while Schweinberg does not confirm any connections between the variables. Regarding the swearing by more highly educated speakers, it was revealed that though L1 speakers seem to curse more often, it is usually behind "closed doors" with friends and family (Dewaele 2017), which findings by Baruch et al. (2016) seem to support, as swearing in high prestige positions was found inappropriate. In Baruch et al. a number of benefits of swearing were highlighted, such as working as a stress relief and a social adhesive. The present paper fills a gap in research by examining a linguistic community that has received very little scholarly attention.

4 Data and methods

4.1 Data

The study utilizes data from the Kaggle data dump[2] website using the chatlogs from the dataset *Dota 2 Matches*. It was compiled using the Opendota platform gathered by using the Steam application programming interface (API), which allows users to collect various types of data from Valve associated titles. The data used in the present study is a part of a larger dataset *Dota 2 Matches,* which includes information on ranked matches played, the playable characters and their frequency of play, match outcomes and chat logs. The data used in the study covers chat

[2] https://www.kaggle.com/devinanzelmo/dota-2-matches?select=chat.csv

interactions from ranked matches which may bear effect on the results of the study. As matches in Dota are divided into two basic types, ranked and unranked, data in this study is based on the former. In Dota 2, playing ranked matches contributes to a player's *matchmaking rating* (MMR), which pits players with similar ratings against each other. Won matches increase one's rating, while lost matches subtract from the overall rating. This means that players are generally motivated to win and are highly averse to losing, which is accounted for in the results of the current study. The raw chat data was downloaded from the Kaggle data dump website. The data was then further divided into two subsets for analysis using Excel, as the corpus analysis program used in the study could not process the original size of the dataset, and therefore had to be divided into smaller datasets to be able to be analysed.

The data were subjected to a preliminary discrimination based on the script and any messages written using either the Cyrillic script or the Hanzi for example, were excluded from the final data set. A rough estimate of the number of messages based on a random selection using Excel's random number feature in languages other than English rests at a 2-3% per 1,000 messages. That is to say that the dataset is a majority English language productions similarly to the study by Märtens et al. (2015: 2), who examine toxicity through the help of text corpora. The final data set contains a total of 3,129,366 tokens.

It should be noted that the information about the sender of a given chat message is highly limited, and no information regarding the region, age, or gender, hours played, frequency of play, or to whom a message was directed at etc. is available. Although demographic data is scarce, some estimates can be made based on the previous literature, as well as the nature of data. As the literature suggests, participants in studies concerning gaming in the medium are mostly male such as in Mattinen and Macey (2018) and others. Furthermore, the male subset involves individuals heavily involved in the virtual environment, as the data stems from ranked matches with personalised player scores, and therefore it is likely that highly achievement-motivated gamers (Tan et al. 2022: 6) are represented in the study. In other words, the data in this study may be more representative of males involved in gaming possessing higher motivation to scale the particular social ladder.

Other than a disappointing lack of demographic data, the dataset used in the study has not received any recent updates, meaning that any findings the study purports are only reflective of the game as it was four years ago. The data analysis of this paper analyses lexical choice as a discourse feature. This means that any conclusion made from the data cannot be representative of any group of users, and hence this study will make broad generalisations about the complexity of discourse in the gaming environment based on any demographic factors, though comparisons with previous studies on swearing may provide interesting parallels.

4.2 Methods

The first objective of the study was to survey the general trends of language use, and a preliminary search in the corpus was conducted and the 10 most frequent types of taboo language were identified. This was done to provide a simple overview of the type of discourse occurring within the discourse. Afterwards another query into the taboo language use was carried out, and a secondary query into the corpus was conducted, extracting the most frequent profanities, swears and general taboo language as defined in section 3**.** The scope of the study presents a number of challenges for the identification of taboo language' due to the very nature of taboo language, as cultural, social, and societal trends may heavily influence the perceptions taboo language, and so the study may contain innate biases in this regard. Prior studies such that of Coats on the spread of profanities in the Nordic countries, for example, employs the use of dictionary annotation of offensive language (2021: 30), which would enable the present study to provide more generally accepted judgements regards any profanities or instances of taboo words for example.

Due to the scope of the study, no generally normative classification such as those of dictionaries are employed in the identification of improper vocabulary, and instead, the necessary vocabulary was selected and annotated based on the author's intuition of the English language, which may be reflected in the results, and is discussed further in the final section of the paper.

The results of the primary query were tabulated using Excel into Tables 2-4 found below. Further queries were conducted into the most frequent items extracted from the first query**,** and their most frequent collocates underwent a similar process on Excel.

The results of the chat corpus were then contrasted with the major English language spoken and written subcorpora of the BNC and the COCA to provide a more holistic view of the discourse, as well as to confirm claims by McEnery and Xiao (2006: 238) who place swearing in the realm of the spoken, in the case of *fuck* and its lemma forms for example. Some essential differences between the BNC and the COCA must however be acknowledged as range of spoken contexts between them differs considerably. Whereas the COCA corpus contains large swathes language from various TV-shows and radio programs, as well as from a number of news casts from channels as the ABC and FOX, the BNC presents a more variable selection of spoken texts from communicative situations in addition to news broadcasts. The spoken BNC, for instance, contains separate sections for pub debates, unscripted and scripted conversations, court hearings, meetings, demonstrations as well as lectures within various disciplines in higher education contexts. It is therefore likely, that due to the higher number of informal communicative contexts of the BNC compared to the COCA, that the COCA corpus likely contains far fewer hits of taboo language, as it may be reasonable to assume that communicational norms regarding non-normative and polite language differ between national broadcasting and conversations proceeding in pubs and bars for example. However, it should be noted that the language of news is not absent from the BNC, even though the range of modes of discourse is wider.

The corpora were accessed, in the case of the BNC through the online interface CQPWeb hosted by Lancaster University (https://cqpweb.lancs.ac.uk)), while the COCA was accessed through a similar online interface through the site English Corpora (https://www.english-corproa.org), and the relevant data was extracted through the search function on both platforms, with relevant restrictions selected. As highlighted by studies such as McEnery and Xiao, swearing is clearly more pronounced in the spoken, and for this reason the corpora and the genres within were the subcorpora of spoken English in both the COCA, and the BNC, and though, for example the

COCA corpus contains numerous subcorpora which contain texts from various spoken channels and contexts such as news, television programs and such, no additional subcorpora were either excluded or specifically highlighted. Instead, for the COCA, this study examines general distributions found in the overall spoken section of the corpus, facilitating comparisons with the chat corpus.

The present study employs corpus methodologies. The use of corpora provides a number of affordances for the researcher, and Leech (2014: 257) notes that corpus methods are currently the most optimal means of studying pragmatics phenomena. Leech asserts that corpus linguistics provides what he terms as 'the God's-eye view" of language, a top-down topological view of the discourse taking place, however as he states, the strain of querying lexical items and patterns such as politeness markers is limited. No method is without its limitations and paralinguistic aspects of language such as prosody, important in mock-politeness for instance, may pose challenges without proper annotation, as is the case of the spoken BNC (Leech 2014: 258), a facet relevant in the present query as well. Although corpus methods provide the present paper with the means of studying individual items in word lists and their collocations, which may display larger trends, the methodology ignores the individual chatting contexts, topics of discussion, referentiality, and much of the nuance vital in determining whether an utterance is intended as impolite, polite, as a face-attack or even as taboo language on its own. Instead, the focus lies on lexical choice.

Furthermore, while corpora allow for swift retrieval of linguistic items, the more abstract components of language, such as irony or sarcasm, may remain from the purview of the analyst (Leech 2014: 259). Moreover, however, as (im)politeness remains a staple of written communication in genres like professional emails, blogs, and other types of CMC, the study of online communication remains tempting (Leech 2014: 260). Although traditional laboratory settings can provide reliable results as suggested by Jay and Janschwitz (2008: 267), corpus methodologies allow for a more direct approach to swearing research. By employing corpus methodologies, the problem of Observer's paradox (Labov 1966: 49), the notion that monitored laboratory settings cause speakers to monitor their speech, and others such as social desirability

bias (Fisher and Katz 2000: 105), i.e., that when asked, informants may answer in socially favourable manner, are neatly circumvented, allowing the present study to cover a somewhat 'hidden' variety of the English language.

The searches into the corpus were conducted using AntConc (Anthony 2023), a corpus tool which provides the more common corpus tools, such as n-grams, word lists, clusters and collocation analysis. The frequencies of each lemma form were retrieved with their most frequent collocates. The results were tabulated and graphically presented using Excel. The preliminary results of the corpus searches are provided in the results section. Though relevant taboo language was identified, a number of items which heavily depend on contextuality of the utterance such as *cancer* were included while terminology relating to religious groups and concepts such as neutral descriptors like *muslim* or *jew* were excluded, in addition to other references like *christ* and *god,* due to their perceived neutral tone. The selection process relied heavily on the lexicon outlined in the literature in section 2.6., placing more emphasis on 'stronger' variants.

Additionally, while the selection process of the relevant vocabulary may be limited, limitations within the reference corpora for the purposes of the study must acknowledged. While the chat corpus exemplifies the linguistic output of a highly select group of speakers, the spoken corpora from the BNC and COCA contain a number of spoken genres ranging from news and TV programs to casual conversations, and as such, any conclusions may only provide superficial conclusions. Finally, misspellings were queried using the asterisk by adding it into the word stem, e.g., *fu*,* which produced variants like *fuk* and *fukk*. Various searches using the wild card and asterisk options on AntConc revealed a number of misspellings and other alternate orthographic forms, which shall be briefly covered as well in reference to items in Table 2 in the following section.

5 Results

Section 4 of the study presents the results of the first corpus study. First the frequencies of the individual lexemes deemed taboo language were tabulated into Table 2, which presents the frequencies of the 10 most frequent types in the corpus, followed by Table 3, which provides an overview of the 25 most frequently appearing types of taboo language. Table 4 provides comparisons between the chat corpus, and the major English language available on the Lancaster corpus database, the British English 2006, and its American English counterpart, American English 2006.

Table 2. Observed and relative frequencies of the ten most frequent items in the corpus per 100,000 words in descending frequency

Rank	Type	Frequency	per 100 000 words
1	lol	481258	15378.8
2	gg	123473	3945.6
3	ez	49235	1573.3
4	i	47936	1531.8
5	you	39782	1271.2
6	u	38547	1231.8
7	wp	33607	1073.9
8	report	30273	967.4
9	a	27473	877.9
10	is	27160	867.9

Table 2 highlights the clear brevity of discourse through the more meaningful or meaning carrying units being acronyms, which the corpus contains the most of: *gg, lol,* and *wp.* The particular acronyms on their own form a jargon within the community as *gg* is the abbreviation of *'good game',* and *wp* the abbreviation of 'well played'. This also highlights a clear set of linguistic norms within the medium, identified as commonplace expressions with a pragmatic

function. due to the sheer frequency of use. Examining the ten most frequent items from Table 2, a clear tendency in the discourse may be observed. Four out of the ten items fall under the category of abbreviations and acronyms in the form of *gg* 'good game', *wp* 'well played', *lol* 'laughing out loud', and *ez,* which in this context seems to present an alternate spelling of *easy* by employing an abbreviated sound-letter correspondence, more akin to its phonological profile.

A brief examination into the concordances of *gg* provides countless examples containing the acronym alone, similarly with *wp,* though more frequently in combination with *gg.* Furthermore, the most frequent collocations for both of these acronyms is the other *gg* for *wp* (1,230,740 hits), and *wp* for *gg.* (336,070 hits). A brief concordance analysis however reveals that all of the acronyms listed here are seemingly used in a repetitious manner, indicating either prolific use as individual messages, or as larger strings of acronyms. Overall, the absence of other lexicon is a notable feature of the data here. Figure 2 highlights this trend, italics added for search term:

(2) ez Ez Ez ez ez ez ez ez ez ez **ez** Ez ez Ez ez ez Ez ez ez ez

When comparing the frequencies of the four items, an observer may notice the remarkably frequent utilization of *lol* and *gg* by the players with *ez* and *wp* at a much lower frequency than the former. With reference to the present study's aims of examining politeness, the stark overuse of *lol*, provides an interesting viewpoint, as *lol* presents frequencies more than three times larger than that of the next most frequent item *gg.* An investigation into the remaining lexical items presents a number of, again, abbreviated, or single syllable linguistic items, with *i, you, u, a, is,* and *report.*

A notable aspect of Table 2 is the presence of a number of personal pronouns as the first and second person pronouns have both been included, with the two forms of the latter both falling in the most frequent word types of the entirety of the corpus. In addition, though *you* abides by the established standard spelling, *i* and *u* are clearly more typical of non-standard discourse since the phonological elements are once more foregrounded as the written form has been abbreviated. Interestingly, though the indefinite article *a* has been included in the most frequent types in the corpus, its counterpart, the definite article *the* has been omitted. Though not

included in the ten most frequent types, the definite article is the 13th most frequent item in the corpus, with a frequency of 23 776, and only marginally less frequent.

Interestingly, although no outright taboo words have been included in the ten most frequent words, *report* has been included. *Report* in the context of MOBAs may refer to the system of moderation by which players may 'report' other players who behave in a disruptive, offensive, or some other manner, the presence of which in this context, provides an indication of the discourse taking place, i.e., the players may perceive the act or idea as relevant or important throughout the medium. Moreover, *report* in this instance may be interpreted as either a noun or a verb, either in the infinitive or in the imperative voice, and so a cursory glance at distributions alone may not reveal whether 'report' functions as an urge towards the reporting of other players, or whether the noun variant for instance serves another function in the discourse.

Moreover, though pragmatic issues may be subject to debate based on corpus frequencies alone, it does highlight the necessity of speedy conveying of information, and the importance of efficient communication. As research in section 2 has highlighted, the nature of the gaming medium clearly situates the activity as swift, reflexive, and constantly changing based on the circumstances of each game. When examining Table 2, this feature is clearly highlighted in the ample instances of abbreviation, and non-standard spelling, which is pronounced in the presence of non-standard spelling variants of the personal pronouns *I* and *you* through the deployment of *i* and *u.* In particular, the first-person pronoun *i,* is clearly more frequent than the other two, though notably, the standard variant of the singular second-person pronoun is clearly present as well in, *you.* It is interesting to note that both of the variants of this pronoun have been included in the most frequent types in the entirety of the 3 million token corpus, which may have a number of communicational implications when examining the tone of the discourse. Firstly, a clearly personal and dialogical communication style is apparent, as players clearly tend to highlight the player themselves, but often refer to other players directly as well. As Xia et al (2019: 501) highlight, there is an element of necessary cooperation which players must meet in order to emerge victorious, which may explain the high frequency of personal pronouns such as

these, as players may highlight their own intentions by placing the focus on themselves linguistically, or alternatively, they may instruct teammates, in this instance in a perhaps direct manner.

Examining Table 2, a clear normative usage of certain acronyms such as *gg, wp,* and *lol* may be identified, as these items constitute the three most frequent items within the data. As politeness may be divided into positive and negative politeness, such as in Brown and Levinson (1987: 317), or pos-politeness, and neg-politeness in Leech (2014), as either the positive traits assigned to an addressee, or the maintenance of boundaries. The acronyms observe in the data seem to on the superficial level, provide a means of maintaining positive face, and building positive affect within the players, either in one's own team, or between the opposing players. *gg* 'good game' acknowledges, according to the theoretical framework by Brown and Levinson (1987: 317) and Leech (2014: 11), the notion that the previous experience has been desirable or enjoyable, i.e., the 'game has been good', and that each participant has acted within the established rules, norms, and expectations, thereby assigning positive face to each of the participants having fulfilled the expected roles for the given match, both one's own team, and the opposing force. The acronyms *wp* 'well played', and *lol* 'laughing out loud' may be characterized similarly, with *wp* being a more targeted vehicle of positive face, as it directly acknowledges an in-game action in a positive manner as desirable. In the case of *lol,* the function may in this instance be conceptualized as a means of building positive affect, as a direct textual parallel of more conventional paralinguistic feature such as the act of smiling or laughing.

In addition, though acronyms such as *gg, wp,* and *lol* may inherently contain an expectation of reciprocity, similar to an expression of gratitude, the number of mitigating features seems low. Furthermore, as the issue of mock-politeness (Culpeper 2013: 9) cannot be ignored in the analysis of the results due to the possibility of irony, sarcasm or other stylistic feature of discourse, the possibility that any negative text-based expression of pragmatic politeness may be null must be accounted for, i.e., any hedging or stylistic negative politeness is likely, due to the previously established bidirectionality, to fall under risk of mock-politeness, or even over-politeness, as outlined by Kerbat-Orecchioni (2013: 20-21). Certainly, no politeness markers

which might be employed to denote neg-politeness by Leech (2014: 11) such as softening mechanisms or hedging can be observed in the data, and so a general lack of negative politeness features may be observed on their own in Table 2, although positive politeness may be interpreted.

Conversely, issues such as negative politeness, as highlighted in the literature may be more difficult to discern, though Table 2 may provide insight. Negative politeness, again as defined by Brown and Levinson, and Leech, underlines the notion of obligation and imposition placed on the individual by the surrounding social environment through pragmatically loaded acts such as expressions of gratitude for example. Table 2 displays a generally high frequency in the usage of the personal pronoun *you* in two various non-standard spelling forms, which may indicate a level of dialogue between the players in a given match, by extension highlighting the possibility that impositions, may take place. The prevalence of the pronoun *you* in its standard form and in its non-standard spelling indicates that players may frequently address each other directly, thus avoiding pragmatic tools such as hedging items or the passive voice, essential in establishing negative politeness (Leech 2014: 312).

5.1 An overview of taboo language in the chat corpus

Table 3. The 23 most frequent types of 'taboo language', observed frequencies and normalised per 100 000 words within the chat corpus

Rank	Type	Frequency	normalised per 100 000
1	fuck	13699	437.756
2	fucking	11290	360.776
3	shit	9828	314.057
4	idiot	3092	98.806
5	retard	2898	92.607
6	bitch	2648	84.618
7	dick	1574	50.298
8	retarded	1539	49.179
9	ass	1535	49.051
10	gay	1250	39.944
11	retards	1056	33.745
12	cunt	1005	32.115
13	fuckin	958	30.613
14	cancer	860	27.482
15	pussy	842	26.906
16	fucker	779	24.893
17	faggot	746	23.839
18	fu	693	22.145
19	rape	655	20.931
20	nigga	602	19.237
22	shitty	564	18.023
23	bitches	499	15.946

Preliminary searches into the use of 'taboo language' in Table 3 within the chat corpus reveal a striking preference for a select number of lexical items far exceeding others in their frequency of

use. The three most common items in the corpus in the realm of taboo language are *fuck, fucking,* and *shit,* with *idiot, retard, bitch,* and *dick* as the next most frequent types within the corpus. In comparison to Table 2, the most frequent forms in the entire corpus, a notable absence, with the singular exception of *fu,* of abbreviation and acronyms. Furthermore, in comparison with Table 2, a marked lack of non-standard spellings such as *i* or *u* may be observed from the data here. Out of the 25 types, only 1 exhibits any form of non-standard spelling in *fuckin* with seemingly little influence from phonological elements which might influence the choice of spelling. Figures 3-5 illustrate the usage of *fuck, fucking* and *shit:*

(3) fuck you nooob fuck you noob shit FUCK YOU NOOBS **fuck** you noobs fuck you nooob

(4) YOU JUST THREW THE **FUCKING** GAME you just throw the game YOU JUST TOO

(5) xd you can just push you know you guys are **shit** you must have thrown YOU WILL FUCKEGN LUCK zeus

When examining the three most frequent types in Table 3, it is clearly visible that the lemma *fuck* clearly possesses a preferred position within the discourse community. Both forms, *fuck* and *fucking,* eclipse their variants by a large margin, indeed, forms such as *fucks, fucked, fucker* and its plural form *fuckers*, do not enjoy nearly as fervent use as the former two. Instead, an alternate spelling of the participial construction *fuckin* seems enjoy more frequent use, followed by the nominalised *fucker,* and the abbreviated *fu*. Similar observations cannot be made for *shit,* with the exception for the adverb *shitty,* though other items in Table 3 do illustrate similar usage as the lemma *retard* may be observed in its plural form as a noun, *retards,* or as an adverb or adjective *retarded,* both of which exhibiting notably more frequent usage than the competing forms *fuckin* and *fuckers.* Similarly, the lemma *bitch* is observed in two of its forms in Table 3, in *bitch* base form, as well as *bitches* in possibly the plural form, assuming the usage pertains to a noun, or to a third-person singular, should the verb be used.

When comparing the distributions of the most frequent instances of taboo language from Table 3, a number of semantic categories may be drafted based on Jay and Jay (2015: 257) and Jay (2009: 153-154) who posit that most swearwords may classified as sexual acts and references thereof, body parts, religious, scatological, references to sexuality, race, or ethnicity, demeaning or pejorative names, or slurs. In the present data of the 25 most frequent words appearing in the corpus. The most frequent categories visible from the table seemingly relate to sexual acts in *fuck, fucking, fuckin,* and though to a lesser degree, *rape,* body parts in *dick, ass, pussy*, scatological references such as *shit* and *shitty,* pejorative or demeaning names such as *idiot, retard, retards, cunt, fucker, bitch,* and *bitches,* as well as slurs targeting sexuality and race such as *nigga* and *faggot.*

Category	**Lexical item**
sexual acts	*fuck, fucking, fuckin,* (*rape*)
body parts	*dick, ass, pussy*
scatological references	*shit, shitty*
pejoratives	*idiot, retard, retards, cunt, fucker, bitch, bitches*
slurs	*nigga, faggot*

Figure 6. Table 3 by lexical category

The list does contain more semantically ambivalent or flexible types which might be categorised in more than one category based on their usage, such as *pussy,* which depending on the surrounding context may be classified as a pejorative insult, an animal, such as in the case of the BNC, with most frequent collocate being *cat,* or a body part. In cases of *fucking,* the usage as an intensifier may be more likely than its lexical meaning due to its higher frequency. Furthermore, some of the words included in the list may on their own even be categorised as neutral, which however become examples of improper language when contrasted with the general environment, such as *gay* and *cancer.* Notably, references to religious themes seem to have been omitted from the list, whereas in the literature, in the Nordics for example by Coats (2021),

religious insults, swears, and curses are highlighted as one of the more frequent categories of lexeme.

When examining the potential impoliteness factors in the data, and by extension the types of face-attacks, which players in the data may perform, a number of issues emerge. As in the case of the language employed by teenagers (Stenström 2017: 168,175), items such as those in Table 3, may be employed to build and maintain positive relations with the co-locutors, and as Culpeper (2013: 9), for example highlights, deciding whether an utterance is deemed polite or impolite is highly context dependent, which may in on their own conflict with notions of face, for instance more starkly. As Culpeper (2013: 135) characterizes, linguistic impoliteness may be achieved through several avenues such by using vocatives, silencers, threats, or negative references of co-locutors, and while face-attacks against the negative face may harder to discern from corpus searches alone, such as those outlined in Table 3, attacks against the positive face may be inferred, although the prevalence of terminology itself on its own may not constitute a face-attack.

Focusing on positive face as defined by Brown and Levinson (1987: 312) as the desired traits an individual may wish to possess, we may identify the desired traits as the opposite of the highlighted traits embodied by the items in Table 3 in addition to factors stemming from the goal-oriented and individualised gameplay and culture surrounding the game itself (Boluk and LeMieux 2017: 213, Korydaka et al. 2020: 1038). As a majority of the items in Table 3 relate to, as outlined in the previous section as well, mental and cognitive abilities or lack thereof (*retard, idiot*), and personal identity or characteristics (*bitch, gay, cunt, fucker, pussy,* and *faggot*), we may notice that the majority of the items listed contain or describe attributes which may be deployed to stigmatise certain features of the addressee. Employing the classifications provided by Spencer-Oatey, we may notice that, out of the 25 most frequent instances of taboo language, most seem to contain a potential challenge to the players' *quality face,* in the form of the previously identified examples*,* and *social identity face* in *retard* and *idiot.*

5.2 Comparing to native data

Comparisons with the major English language corpora provide interesting results, and a drastic difference in the distributions observed in the chat corpus. The results of the comparative efforts have been summarised in Table 4. However, as briefly discussed in section 3.2, it must be noted that the range of contexts within the COCA and the BNC and the norms that motivate language use in these domains are markedly different. In particular, the COCA corpus contains nearly exclusively spoken conversations within public forums on televised broadcasts, which subsequently diminishes the likelihood of taboo language being uttered by speakers. This range of registers may also be relfected in the frequency of taboo language when comparing the COCA to the BNC, as the results already indicate higher frequency of lexical items pertaining to taboo language, likely as a result of the inclusion of categories such as 'unscipted conversations' and 'pub debates'. The results demonstrate clearly demonstrate a major difference between standard English and non-standard English.

Table 4. Normalised frequencies of items from Table 3 across corpora

TYPE	chat corpus	BNC	COCA
fuck	437.76	4.83	0.01
fucking	360.78	18.04	0.00
shit	314.06	5.85	0.05
idiot	98.81	0.73	0.56
retard	92.61	0.00	0.03
bitch	84.62	1.15	0.42
dick	50.30	1.68	5.56
retarded	49.18	0.09	0.28
ass	49.05	0.17	0.58
gay	39.94	0.81	6.44
retards	33.74	0.00	0.01
cunt	32.12	0.79	0.00
fuckin	30.61	0.00	0.00
cancer	27.48	1.60	9.47
pussy	26.91	0.78	0.03
fucker	24.89	0.18	0.00
faggot	23.84	0.03	0.03
fu	22.15	0.00	0.00
rape	20.93	0.88	3.77
nigga	19.24	0.00	0.00
shitty	18.02	0.18	0.00
bitches	15.95	0.08	0.06

When comparing the corpora based on the frequencies presented in Table 4, it becomes clear that the values in the chat corpus are considerably higher than in either of the reference corpora. Based on a cursory observation of the tables, it is evident that players of Dota produce more of these particular swears than either the British or the Americans, which is applicable to each of the items identified in Table 3. In fact, no item, even the most frequent swears identified

in the literature by the likes of McEnery and Xiao (2006), reaches similar levels of use, when the values are normalised to 100 000 words. An observation into the frequencies does, however, provide some interesting details, as in some instances, the American corpus far exceeds that of the chat corpus and the British one. Namely the words *dick, gay, cancer* and *rape* appear far more frequently in absolute terms than they do in the chat corpus, though when compared to the normalised frequencies, the chat corpus still far exceeds the level of use observed in the reference corpora. As iterated in section 3, the reference corpora contain linguistic material from a wide range of spoken genres, which determine the acceptability of taboo words according to their own criteria, as news language is less likely to employ taboo language than a casual conversation for example, and may also employ a more descriptive tone in general rather than a dialogical or confrontational.

Comparing the normalised frequencies in Table 4, it is clear that a taboo language is, by comparison, a staple of the medium, as no single item within the two reference corpora neither exceeded the frequencies present in the chat corpus nor were able to parallel the frequencies put forward by the chat corpus. Additional comparisons between the two reference corpora highlight the closer resemblance of the chat corpus and the British corpus, however. It was found that the three most frequent items from Table 3, the taboo language employed by chatters in DotA, were the closest in frequency to the BNC corpus, though admittedly the values provided by the chatters are still considerably higher. Comparisons with the American speakers provide yet more resemblance, and while a similar disparity in distributions between the chat corpus and the American corpus is observable, there is additionally little overlap between the British and the American corpus. Overall, the American corpus contains very little swearing, even when compared to BNC, as the British speakers use vocabulary more similar to the chatters' with *fuck, fucking,* and *shit* more frequently, while the usage of these words is nearly absent in the American corpus. Interestingly, there is noticeable overlap in case of the words *cancer, gay, dick* and *rape,* which are noticeably more frequent in the American corpus, than in the British, yet less frequent in the chat corpus when observing the normalised values alone.

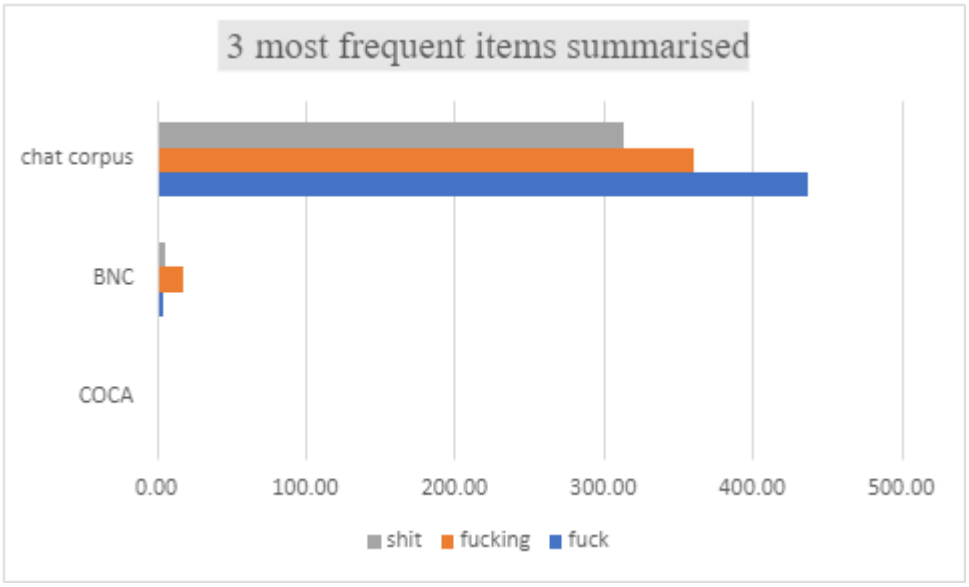

Figure 7. Three most frequent items from Table 4 summarised, *shit, fucking,* and *fuck*

Interestingly, the chat corpus seems to bear the starkest resemblance to the British corpus, as the most frequent items employed by the players are also those most frequently employed by the British, when comparing the items in the chat corpus to those of the other datasets. Based on the distributions here it seems that most of the swears employed by players of Dota are hardly used by either of reference groups, and instead only four of the swears prevalent in the chat corpus overlap with the native speaker corpora, i.e., the British corpus seems to favour *fuck, fucking,* and *shit,* with more minor usage of *dick* and *bitch.* By contrast, the COCA, and the chat corpus would seem to have very little in common, when compared on the most frequently appearing swear words in the chat corpus, and even when compared against the British corpus. Similarly, the lexemes with which the COCA differs from the BNC are the ones which differentiate COCA from the chat corpus, and it seems that overall, the American speakers may produce fewer swears than do the players or the British speakers of English.

Overall, when examining the most frequent vocabulary items used to deploy taboo language, the present study illustrates a mixed behaviour of use, highlighting, on the one hand features of the general BrE, and on the other hand, features more distinctly AmE. As it was found that the most frequent curse or swear words within the players of Dota 2 are in fact, *fuck, fucking,* and *shit.* This provides some indication as to the applicability of their results and do align to some degree with

the present study. When comparing the findings in the present study to those of McEnery and Xiao (2004: 236) we may observe clear corroborative findings. In their study *fucking* was found to be more frequent than *fuck* (McEnery and Xiao 2004: 236), a finding which was confirmed in the present study as well, as chatters in the DotA 2 gaming context seem to favour these words, and moreover, their frequencies far exceed those of the remaining items in Table 3. *fuck* and *shit* in particular not only constitute an important aspect of British profanity (Love 2021:750-751), *fuck* specifically is commonly exploited by teenager speak (Martinez and Petrejo 2012: 782).

It is worth noting that, as the study by Vatvedt (2019: 101) illustrates, especially the word *fuck* has been incorporated into various L2 varieties in the Nordics, and Russia for instance, having been assimilated and adapted in its phonological, orthographical, and grammatical elements to various degrees, while retaining its stylistic and register-specific character as a feature of informal spoken registers. It may be the case then, that due to the holistic assimilation into L2 varieties, these particular forms may be deployed more readily in speech than culture-sensitive alternatives such as religious and cultural concepts, or taboo language specific to gender, as gender norms vary across groups.

When comparing the findings of the chat corpus as well as those of the British L1 speakers, we may note that the seemed to contain far fewer instances of the three types more frequently observed in the two other corpora. As has been noted in the literature, norms regards swearing or taboo language vary culturally, and British speakers may, for instance, find it more appropriate to deploy taboo vocabulary in the workplace, while conversely, American speakers may find similar use inappropriate (Dewaele 2015: 323-324). It is therefore reasonable to believe that similar normative stances might be reflected in the results of the paper, reflecting broader cultural stances or cultural influences more generally. Although cultural stances may apply, perceptions of offensiveness may vary between speaker groups as well, as Janschewitz (2008: 269-270) for instance notes, the offensiveness of taboo words may be deemed higher by L1 speakers. McEnery and Xiao (2004: 246) further characterise the swearing in their study by noting that the use of swear words was the most frequent with younger speakers aged 15-24-

and 24–34-year-olds, which again demonstrates a correspondence between studies such as Mattinen and Macey (2018).

Additionally, the gender of the speakers and a target male audience seem to facilitate the production of taboo language in the literature, and while the demographic details of the data set on DotA 2 are not available, we may compare the literature against the findings, and as highlighted in the literature. As most participants in studies focusing on gaming, as highlighted in section 2, are men, and as men are more likely to produce toxic language overall, these comparisons become intriguing. Functionally, *fucking* is most often characterised as an emphasiser, or a booster, while *fuck* is more frequently a general curse word, or an expletive (McEnery and Xiao 2004: 258). Schweinberger's findings situate the use of *fuck* into an emphatic role and emphasise the same-sex interlocutor effect, highlighting the previous notion of target-audience within swearing frequencies (Schweinberger 2018: 13). He notes that the regional and gender-based variation may be a way to construct social ties with speakers of a certain variety.

Overall, the data observed in the previous study clearly supports the finding established in the literature that men in general, produce more taboo language (Coats 2021: 38-39, Schweinberger 2018: 11, McEnery and Xiao 2004: 240,248, and Love 2021: 751-752), further amplified by the assumption that the chatters in DotA 2 localise their linguistic productions towards a male audience, as male target-audience has been identified as an important factor in the frequency of taboo language as well by previous research (McEnery and Xiao 2004: 250), as male a perceived male audience is likely to incite greater production of taboo vocabulary, such as in the case of *fucked*. The extremely high frequencies observed in the present Dota corpus may be facilitated by similar factors, though further qualitative and quantitative examination between corpora is needed.

6 Discussion

This study has examined the issue of taboo language within the context of DotA 2 through the use of a chat corpus of messages extracted from ranked matches. Several striking findings have made which characterise the discourse within the particular speaker community and may be briefly summarised in terms of the relative brevity of expression. The main purpose of the present study has been to examine and describe the distributions of bad or taboo language and provide comparisons to a wider range of discourses through the use of the British National Corpus or BNC, and the COCA. While the comparisons extracted throughout these corpora largely ignore the wider borders on genre, and instead focuses on the larger macro-scale frequencies to provide an overview of the relevant findings, as well as provide a 'real-world' comparison to the virtual realm, there are findings which underline the unique features of the medium of online gaming in this regard.

The results show that speakers within the chat corpus employ considerably more swears, curse words, and other types of taboo language, than what is observed in corpora that contain standard American English, or British English. Notably, speakers seem to favour three particular items when producing taboo language, namely *fuck, fucking,* and *shit,* which align with the usage observed in the BNC but which are not present in the COCA corpus. Further analysis indicates that aside from the previous lexical items, no one item in either corpus reaches similar levels of use in this study. This highlights the fact that online modes of discourse such as the one examined in the present study, abide by a completely separate set of norms than comparable sets of standard English. In addition to the frequencies between the standard English corpora and the chat corpus, comparisons between Table 2 and Table 3 highlight a clear absence of non-standard spelling within the range of taboo language used by players, which demonstrates a degree of paradoxicality. On the one hand, players favour swift, condensed language, while on the other hand all observed types of taboo language, with the exception of *fu* and *fuckin,* demonstrate little to no utilization of non-standard spelling or acronyms, unlike in Table 2. This seems to indicate a clear deliberateness and would suggest that the taboo language occurring

within is a clear example of propositional swearing, used to fill a social function (Janschewitz 2008: 269-270).

As the literature on cursing and taboo language illustrates, and which is reflected in the present study, most instances of taboo language employed by speakers tend to fall under various categories, including religion, sexual acts and sexual references, scatological and excretory references, deviant behaviours, pejoratives and insulting names, as well as others such as ethnicity, race or sexuality, as outlined by the likes of Love (2021: 742), and Jay (2009: 153). In the present study, the present categories were all identified with the most frequent categories being the sexual references such as *fuck* and *fucking,* which were the most frequent, scatological references in *shit,* as well as a number of pejorative names denoting personal identity or characteristics such as *retard, idiot, faggot* and so on.

Interestingly, as Jay and Jay (2015: 254-255), as well as Rosenberg et al. (2016: 309) note, a more closed set of lexical items used by speakers was identified, with marked overuse of the three most frequent items in the data, as *fuck, fucking,* and *shit,* seemed to enjoy noticeable overuse by the chatters. In addition to the intensifiers situated in the three most frequent items in Table 3, there is a marked prevalence of the categories of lexemes denoting personal identity, and in particular words denoting mental disability, lack of intellect, or insufficient cognitive capabilities, such as *retard* and its plural form, *idiot* are situated in the likes of more frequently used items. Furthermore, there is an utter lack of references to religious concepts such as Hell or God, which may be stem from a semantic shift regards the meaning and pragmatic power of the particular lexicon, though favoured by those in Nordics, Stenström notes that speakers of may perceive the tabooness religious lexicon lesser due to semantic bleaching (2017: 168, 175). Noticeably, feminine coded vocabulary such as *bitch, cunt,* and *pussy* are also used frequently by the chatters, with *bitch* maintaining the highest comparative frequency, and occupying a similar range with *retard* and *idiot,* indicating perhaps a similar pattern of use.

Moreover, when comparing the results of the present study to those by Kwak and Blackburn (2015: 213) on swearing within *League of Legends,* a similar MOBA to the present inquiry, similar

cursing inventories can be identified, though within the 23 most frequently employed swearwords within the present data, only some are included in the study by Kwak and Blackburn. In LoL, some of the more frequent instances of taboo language included, which were identified in the present data as well, *retard,* and non-standard spellings of *fucking* such as /fuckign/, /fking/ and /fukin/. Interestingly, and similarly to the current study, the intensifier *fucking* seems to collocate with *report, noob, retard,* as well as in some constructions which were not identified in the present investigation such as *report fucking* or *play fucking* (Kwak and Blackburn 2015: 213). When contrasting the lexical inventories of the studies by McInroy and Mishna (2017: 602-603) and Blackburn, a notable absence of vocabulary such as *pwn, noob,* and *newb* from the study by McInroy and Mishna, as well as lexical items and combinations such as *pussy ass, nooob, pathetic,* as well as ethnic and racial slurs are mostly omitted, with the exception of *nigga* in the present data. This comparison underscores the object of toxic language, i.e., players employ clearly more numerous instances of pejoratives targeted at a perceived lack of cognitive ability or intellect, and less frequently their experience within the game, as terminology similar McInroy and Misha's study was completely absent from Table 3. When examined through the framework of impoliteness research explicated in the third research question for the paper, the present study provides a number of findings, though the applicability of the framework such as those outlined in Leech (2014) or Brown and Levinson (1987) may be limited due to the nature of the medium as a mainly a text based communication channel. As the present study is only able to investigate the distribution, as well as the qualitative aspects of virtual communication through the text format players engage in, issues such as prosody, gestures, and other general paralinguistic elements cannot be commented on. However, there are a number of findings which may be examined regardless, such as the taboo language in Table 3, and their relevance in face work.

7 Conclusion

Through the use of a corpus of chat messages, the present paper has examined the frequency of the 25 most frequent types of taboo language in produced by players during ranked matches of DotA 2 . It has also compared their frequencies to two major English language corpora, the British National Corpus (BNC), and the Corpus of Contemporary American English (COCA). The present study has highlighted a marked overuse of taboo language in the chat corpus for all types presented in Table 3, far exceeding either of the two observed reference corpora. This study has additionally managed to find overlap with previous literature, as the most frequent items in Table 3 *fuck, fucking* and *shit,* have been identified as the most frequent British English for example by previous literature. Although the distributions observed within the chat corpus are multiple times larger than either of the compared datasets, some similarities between the corpora were identified. Namely, the observed frequencies of the types *fuck, fucking,* and *shit,* neatly correspond in their frequency of use to the distributions observed in the British corpus. It was observed that, by comparison, the American corpus contains far fewer instances of swearing overall, however some items, such as *gay, cancer, nigga,* and *dick* seem to have been employed more frequently in absolute terms than in either the chat corpus, or the BNC. Regardless of the frequencies of the previous items, the chat corpus demonstrates far more frequent use of all compared items in normalised frequencies, reaching nearly 400 times more frequent use in the case *fuck* and *fucking* for example. In addition, more general characterisations of the discourse were derived from a brief analysis of the most common types in the chat corpus, through which it was demonstrated that players of DotA 2 prefer curt exchanges, and readily condense information into acronyms, and cut linguistic corners by employing non-standard spelling readily. This analysis has been able identify some discourse specific jargon bordering on pragmatic markers as the analysis in section 5 on politeness and face has been able demonstrate, noting that acronyms such as *gg* contain inherent expectations of reciprocity.

This study and the analysis conducted has been able to extend the line of research on impoliteness by employing a novel form of data, employing compilations of chat messages of a

popular video game, and thus avoiding many of the pit falls of linguistic research such as the Observer's paradox. Furthermore, the data of the present study is comprised of naturally occurring, unmonitored, user-generated data, and thus illustrates the degree to which speakers may feel comfortable extending their use of non-standard or stigmatised lexical items and patterns of use. It is highly likely, based on the results of the study, that when combined with total anonymity, transient and changing social bonds, and mostly text-based channels of interaction, speakers' linguistic output may drastically change. However, it must be noted that the present study is likely to examine a highly specific group of speakers, and to draw any parallels with larger trends may be difficult.

As previous analyses of the medium of online gaming, DotA 2 and the MOBA genre more generally, have received only a limited number of papers, mostly examining the technical aspects of online gaming, consumer habits, and so on, this paper serves as an important stepping stone in establishing a linguistic perspective, and as such basic information regards the particular variety have been established neatly. Similarly, as the results of the paper provide results markedly larger frequencies of taboo language than either of the reference corpora for instance, this paper has brought light onto a neglected language variety and provides an interesting focal point for future research. This paper is, to my knowledge, one of the few papers attempting to apply pragmatic theory, i.e., face and politeness to online gaming, and certainly to the MOBA-genre. In this manner, the paper has been able to provide a novel approach towards face research in this aspect by combining the discipline with corpus methodologies and managed to provide credible results.

Although the paper has been able to provide results on the topic and has broken new grounds in the study of online gaming, a number of issues and limitations might influence its main findings and should be accounted for in future research. Firstly, as taboo language is highly dependent on the cultural, social, as well as local norms, it is inevitable that any analysis is subject to researcher bias. In particular, the identification of taboo language requires more rigorous methodology, in order to avoid including neutral, or otherwise irrelevant data points, as may be the case with the present study's findings regarding Table 4. Table 3, for example, presents the

lexical items *gay,* and *cancer,* which both possess a multiplicity of semantic meanings, which only become taboo in specific contexts. As such approaches such as those of Coats (2021), i.e., the utilization of annotated dictionaries, should be developed to eliminate any particular researcher bias from the equation. Furthermore, while the quantitative results highlighted by this paper provide indication of the discourse proceeding within, more qualitative analysis is needed in to establish more detailed and fine-grained analysis, as the present study is only able to provide information about the general trends inside. Qualitative analysis should be extended in particular to highlight any possibility that acronyms like *gg* for instance are actually pragmatically meaningful, for example. As Table 2 may indicate, acronyms are a staple of the gaming medium, and some combinations are clearly more frequent than others (*gg ez*).

This study has been able to highlight some of the discourse features of online video game chat with a focus on taboo language, and while a number limitations plague the design study of the study and the data, the study has been able to discover novel aspects of a largely neglected speaker community. Future research should focus more attention on the qualitative aspects of the medium, since, as this study has demonstrated, a different set of norms clearly dictates the ongoing discourse. Factors such as anonymity, and fast paced actions seem to exert a non-insignificant influence on the discourse, and a variety of non-standard features are present due to these factors. As this study has highlighted, characterisations of non-standard language features may present an intriguing field of research, as acronyms and abbreviations were rife in the present study.

In addition, further qualitative analysis is needed to establish the exact reasons for the exceedingly high frequency of swear words and taboo language, as there are clear differences between the spoken corpora observed in the study, and the chat corpus. While the present study has been able to bring to light some of the issues surrounding politeness and face, further research is required to determine the extent of mock-politeness and more pressing socially mediated relations phenomena, i.e., how do the players situate themselves within the transient and temporary discourse environment, and how do they negotiate politeness within this space. Online gaming presents a unique research opportunity, with ample opportunity for novel

findings such as the ones presented by the present paper, with little research on the linguistic perspective, and as such more attention is required to understand this marginalised mode of discourse.

References

Akhmedov, Kodirjon and Anh Huy, Phan 2021. “Machine learning models for DOTA 2 outcomes prediction”. *arXiv.Org.*

Anthony, Laurence. 2023. AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software.

Ballard, Mary E., and Welch, Kelly M. 2017. “Virtual warfare: cyberbullying and cyber-victimization in MMOG play. *Games and culture,* Vol. 12(5), pp. 466-491.

Baruch, Yehuda, Prouska, Rea, Ollier-Malaterre, Arianne and Bunk, Jennifer 2017. “Swearing at work: the mixed outcomes of profanity”. *Journal of managerial psychology,* Vol. 32 (2), pp. 149-162.

Boluk, Stephanie, and LeMieux, Patrick 2017. *Metagaming : Playing, Competing, Spectating, Cheating, Trading, Making, and Breaking Videogames*, University of Minnesota Press.

Brown, Penelope and Levinson, Stephen C. 1987. “Politeness: some universals in language usage”. In Adam Jaworski and Nikolas Coupland (eds.), *The Discourse Reader,* 2006, London: Routledge.

Bu3ny. “Game map 7.33”. *Dota 2 Wiki,* available online at: https://dota2.fandom.com/wiki/Map .
Coats, Steven 2021. “‘Bad language’ in the Nordics: profanity and gender in a social media corpus”. *Acta linguistica hafniensia,* Vol. 53(1), pp. 22-57.

Cook, Chrissy, Schaafsma, Juliëtte, and Antheunis, Marjolijn 2017. “Under the bridge: an in-depth examination of online trolling in the gaming context”. *New media & society,* Vol.20(9), pp. 3323-3340.

Culpeper, Jonathan 2011. *Impoliteness: Using Language to Cause Offense.* Cambridge University Press.

Culpeper, Jonathan 2013. "Impoliteness: questions and answers". In (eds.) Denis Jamet, Manuel Jobert *Aspects of Linguistic Impoliteness,* Cambridge Scholars Publishing.

Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA).* Available online at https://www.english-corpora.org/coca/.

De Mesquita Neto, Alvino, Joaquim, and Becker, Karin 2018."Relating conversational topics and toxic behaviour effects in a MOBA game". *Entertainment computing,* Vol. 26, pp. 10-29.

Devin, Anzelmo, 2020. *Dota 2 Matches: Explore Behavior and Predict Match Outcomes.* Available online at https://www.kaggle.com/devinanzelmo/dota-2-matches?select=chat.csv.

Dewaele, Jean-Marc 2015. "British 'bollocks' versus American 'jerk': Do native British English speakers swear more – or differently – compared to American English speakers?". *Applied linguistics review,* Vol. 6 (3), pp. 309-339.

Dewaele Jean-Marc 2017. "Self-reported frequency of swearing in English: do situational, psychological and sociobiographical variables have similar effects on first and foreign language users?". *Journal of Multilingual and Multicultural Development,* 38(4), pp. 330-345.

Fisher, Robert, J. and Katz, James, E. 2000. "Social-desirability bias and the validity of self-reported values". *Psychology and marketing,* Vol. 17(2), pp. 105-120.

Fox, Jesse and Tang, Wai Yen 2017. "Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies". *New media & society,* Vol. 19(8), pp. 1290-1307.

Goffman, Erving 1967. *Interaction Ritual: Essays on Face-to-Face Behavior.* New York: Doubleday.
Hardie, Andrew. “British National Corpus (XML-edition)”, CQP-Web, Lancaster University.

Hilvert-Bruce, Zorah, and Neill, James, T. 2019. “I’m just trolling: the role of normative beliefs in aggressive behaviour in online gaming”. *Computers in human behaviour,* Vol. 10, pp. 303-311.

Ismangil, Milan 2019. “Subversive nationalism through memes: a Dota 2 case study”. *Studies in ethnicity and nationalism,* Vol. 19(2), pp.227-245.

Jane, Emma, A. 2015. “Flaming? What flaming? The pitfalls and potentials of researching online hostility”. *Ethics and information technology,* Vol.17(1), pp. 65-87.

Jarret, Josh, Brock, Tom, and Johson, Mark 2021. “Gaming the gift: the affective economy of League of Legends ‘fair’ free-to-play model”. *Journal of consumer culture,* Vol.21(1), pp. 102-119.

Jay, Kirstin, L., and Jay, Timothy, B. 2015. “Taboo word fluency and knowledge of slurs and general pejoratives: deconstructing the poverty-of-vocabulary myth”. *Language sciences,* Vol.52, pp.521-259.

Jay, Timothy 2009. “The Utility and Ubiquity of Taboo Words”. *Perspective on psychological science,* Vol. 4(2), pp.153-161.

Jay, Timothy and Janschewitz, Kristin 2008. “The pragmatics of swearing”. *Journal of Politeness Research,* 4, pp.267-288.

Katona, Adam, Spick, Ryan, Hodge, Victoria, J., Demediuk, Simon, Block, Florian, Drachen, Anders, and Walker, James, Alfred 2019. “Time to die: death prediction in Dota 2 using deep learning”. *IEEE Conference on Games (CoG),* pp.1-8.

Kerbrat-Orecchioni, Catherine 2013. “Politeness, impoliteness, non-politeness, “polirudeness”: the case of political TV debates”. In (eds.) Denis Jamet and Manuel Jobert *Aspects of Linguistic Impoliteness,* Cambridge Scholars Publishing.

Kordyaka, Bastian, Jahn, Katharina, and Niehaves, Bjoern. 2020. “Towards a unified theory of toxic behaviour in video games”. *Internet research* Vol.30 (40), pp. 1081-1102.

Kwak, Haewoon, and Blackburn, Jeremy 2015. “Linguistic analysis of toxic behavior in an online video game”. In *Lecture Notes in Computer Science* (including subseries *Lecture notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), Vol. 8852, pp. 209-217.

Labov, William 1966. “The Social Stratification of (r) in New York City Department Stores”. In Harold B. Allen and Michael D. Linn (eds.) *Dialect and Language Variation.* Academic Press Inc.
Leech, Geoffrey, N. 2014. *The Pragmatics of Politeness.* Oxford University Press.

Lee, Matthew, Johnson, Daniel, Tanjitpiyanond, Porntida and Winfred, Louis. R. 2022. “It’s habit not toxicity, driving hours spent in DOTA 2”. *Entertainment computing*. Vol. 41, pp.
Love, Robbie 2021. “Swearing in informal spoken English: 1990s-2010s”. *Text & talk,* Vol. 41(5), pp. 739-762.

Martinez Ignacio M. Palacios and Pertejo Paloma Núñez 2012. “He’s absolutely massive. It’s a super day. Madonna, she is a wicked singer. Youth language and intensification: a corpus-based study”. *Text & Talk* 32(6), pp. 773-796.

Mattinen Topias and Macey Joseph 2018. “Online Abuse and Age in Dota 2”. *Proceedings of the 22nd International Academic Mindtrek Conference*, 2018-10-10, pp. 69-78.

McEnery, Anthony, and Xiao, Zhonghua 2004. “Swearing in modern British English: the case of fuck in the BNC”. *Language and literature,* Vol. 13(3), pp. 235-268.

McInroy, Lauren B., and Mishna, Faye 2017. “Cyberbullying on online gaming platforms for children and youth”. *Child & adolescent social work journal,* Vol. 34(6), pp. 597-607.

Monge, C., K., and O’Brien, T., C. 2022. “Effects of individual toxic behaviour on team performance in League of Legends”. *Media psychology,* Vol. 25(1), pp. 82-105.

Mora-Cantallops, Marcal and Sicilia, Miguel-Ángel 2018. “MOBA games: A literature review”. *Entertainment computing,* Vol. 26. pp. 128-138.

Murnion, Shane, Buchanan, William, J., Smales, Adrian and Russel, Gordon 2018. “Machine learning and semantic analysis of in-game chat for cyberbullying”. *Computers and security,* Vol. 76, pp. 197-213.

Märtens, Marcus, Shen, Siqi, Iosup, Alex, and Kuipers, Fernando 2015. “Toxicity detection in multiplayer online games”. In *2015 International Workshop on Network and Systems Support for Games (NetGames).*

Ortiz, Stephanie, M. 2019. ““You can say I got desensitized to it”: How men of color cope with everyday racism in online gaming”. *Sociological perspectives,* Vol. 62(4), pp. 572-588.

Ratan, Rabindra, A., Taylor, Nicholas, Hogan, Jameson, Kennedy, Tracy, and Williams, Dmitri 2015. “Stand by your man: an examination of gender disparity in League of Legends”. *Games and culture,* Vol.10(5), pp. 438-462.

Rosenberg, Patricia, Sikström, Sverker, and Garcia, Danilo 2017. ”The A(ffective) B(ehavioral) C(ognitive) of taboo words in natural language: the relationship between taboo words’ intensity and frequency”. *Journal of Language and Social Psychology,* Vol. 36(3), pp. 306-320.

Röhlcke, Sebastian, Bäcklund, Christian, Sörman, Daniel, Eriksson, Jonsson, Bert, and Kotozaki, Yuka 2018. “Time on task matters most in video game expertise”. *PloS one,* Vol. 13(10), pp.

Schweinberger Martin 2017. “Swearing in Irish English – A corpus-based quantitative analysis of the sociolinguistics of swearing”. *Lingua,* 209, pp. 1-20.

Sengün, Sercan, Santos, Joao., M., Salminen, Joni, Jung, Soon-gyo, and Jansen, Bernard, J. 2022. “Do players communicate differently depending on the champion played? Exploring the Proteus effect in League of Legends”. *Technological forecasting & social change,* Vol. 177, pp.

Spencer-Oatey, Helen 2002. “Managing rapport in talk: Using rapport sensitive incidents to explore the motivational concerns underlying the management of relations”. *Journal of Pragmatics,* Vol.34 (5), pp. 529-545.

Stenström, Anna-Brita 2017. “Swearing in English and Spanish teenage talk”. In (eds.) Kristyn Beers Fägersten and Karyn Stapleton, *Advances in Swearing research: new languages and contexts*. Amsterdam: John Benjamins Publishing Company, pp. 157-181.

Stokes, Benjamin, and Williams, Dhimitri 2018. “Gamers who protest: small-group play and social resources for civic action”. *Games and culture,* Vol.13(4), pp.327-348.

Suler, John 2004. “The online disinhibition effect”. *Cyberpsychology & behavior,* Vol.7(3), pp.321-326.

Sulpizio, Simone, Toti, Michelle, Del Maschio, Nicola, Costa, Alberta, Fedelli, Davide, Remon, Job, and Abutalebi, Jubin 2019. “Are you really cursing? Neutral processing of taboo words in native and foreign language.” *Brain and Language* 194, pp. 84-92.

Sweeney, Kevin, Tuttle, M., H., and Berg, Douglas, M. 2021. “Esports gambling: market structure and biases”. *Games and culture,* Vol.16(1), pp. 65-91.

Tan, Wee-Kheng, and Chen, Lu-Ming 2022. “That’s not my fault: excuses given by layers exhibiting in-game intra-team aggressive behavior in online games”. *Computers in human behaviour,* Vol.127, pp.

Tang, Wai Yen, and Fox, Jesse 2016. “Men’s harassment behavior in online video games: personality traits and game factors”. *Aggressive behavior,* Vol.42, pp.513-521.

Teng, Ching-I, Tseng, Fan-Chen, Chen, Ye-Sho, and Wu, Soushan 2012. “Online gaming misbehaviours and their adverse impact on other gamers”. *Emerald Group Publishing Limited,* Vol. 36 (3), pp. 342-358.

Vatvedt Fjeld, Ruth, E., Kristiansen, Elsa, Rathje, Marianne, Oskarsson, Veturlidi, Kostannovskaia, Natalia, Gill, Inyat, and Menuta, Fekede 2019. “The worldwide use and meaning of the f-word”. *Intercultural pragmatics,* Vol.16(1), pp.85.111.

Wells, Garrison, Romhanyi, Agnes, Reitman, Jason, G., Gardner, Reginald, Squire, Kurt, and Steinkuehler, Constance 2023. “Right-wing extremism in mainstream games: a review of the literature”. *Games and culture,* pp. 1-24.

Xia, Bang, Wang, Huiwen, and Zhou, Ronggang 2019. “What Contributes to Success in MOBA Games? An Emprical Study of Defense of the Ancients 2”. *Games and Culture,* Vol 14. (5), pp. 498-522.

Zhang, Wenjie, Zou, Lei, Maamar, Zakaria, and Chen, Liu 2021. “MOBA game analysis system based on neural networks”. In *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), Vol.13081, pp. 518-526.