Topics in Linguistics (2026), 27(1), pp. 103–125
10.17846/topling-2026-0005

Gaming language stratification: Genre-based vocabulary specialization

Martin Kažimír, Patrícia Kubišová*
Constantine the Philosopher University in Nitra, Slovakia

Key words

gaming language
lexical comprehension
semantic components
inter-rater agreement
mixed ANOVA

Abstract

This pilot exploratory study investigates whether English-language gaming jargon is a single, homogeneous register, or is stratified into partially distinct, genre-specific jargons. We collected comprehension judgments for lexical items from six game genres using a semantic-component questionnaire and a canonical "answer key" developed with domain consultants. Agreement among respondents was quantified using Fleiss's κ to assess within-group consistency and Cohen's κ to compare group consensus to the canonical key. Complementary mixed (repeated-measures × between-subjects) ANOVAs tested comprehension variance by genre and by reported genre experience. Results show systematic differences in comprehension across genres and a consistent pattern of higher mean comprehension among self-identified players for four genres; two genres did not show a player advantage. The canonical key was experimentally validated by player consensus. Given the exploratory design, conclusions are tentative; findings nonetheless provide preliminary evidence that gaming language displays genre-based stratification and warrant follow-up studies.

1. Introduction

Gaming language is usually treated as a single, more or less homogeneous language stratum. Even though Ensslin points out already in 2012: "it is near impossible to speak of a language of videogames in the singular […] some game genres differ so radically in their textual and ludic make-up" (Ensslin, 2012, p. 42), virtually no attention has been paid to the actual stratification of gaming jargon. While this omission in research on gaming language is understandable – the fast-paced evolution of the medium in question makes it difficult to draw any long-lasting conclusions of a more abstract kind – we believe that the stratification of gaming language significantly shapes both gamer identities and the interactions between people participating in gaming discourses.

* Corresponding author

Cite this paper: Kažimír, M, & Kubišová, P. (2026). Gaming language stratification: Genre-based vocabulary specialization. *Topics in Linguistics, 27*(1), 103–125. https://doi.org/10.17846/topling-2026-0005

This pilot study aims to provide exploratory observations about the degree of English-language gaming jargon stratification that will serve as a foundation for more robust approaches to observing how jargons of individual gamer groups interact with each other, and to verify the operationality of a specific questionnaire design intended as a tool for vocabulary comprehension estimation. The choice of English is motivated by the simple fact that it is the *lingua franca* of gaming discourses. It is the language in which most games are published, most game-related content is created, and which most gamers, regardless of their native languages, use to communicate with the larger communities. The choice of lexicological focus is motivated by the facts that 1) gaming language is primarily a lexically conditioned jargon, with lexical specialization in particular recognized as the factor shaping gamer identity (Bussmann, 1996, p. 245); and that 2) among language levels, vocabulary is the one most immediately accessible to self-reflection.

2. Overview

2.1 Sociocultural relevance of video games and gaming language and research areas of interest

Games in general, and video games in particular, are not only entertainment; they are shaped by, and, in turn, shape societal values, cultural trends and human behaviour. For instance, there is a growing trend among game developers to produce games that "aim to increase the public's awareness of a wide range of pressing global social issues, including poverty, cancer, child abuse, gender dysphoria, and refugee crisis" (Shliakhovchuk, 2024, p. 1). Esnaola and San Martín (2004, apud Cerezo-Pizarro et al., 2023) claim that the relationship between a video game and culture exists due to a game's capacity to disseminate and transmit knowledge, beliefs, ideas and values, but at the same time, to generate new cultural representations that are visible within the game rules and dynamic structures.

The video game industry is attracting increasing academic attention, partially because it is an area of consistent increase in sales and use. For 2025, the game market is forecast to reach USD 522.46bn in sales (Statista, 2025) and more than three billion active video game players (Duarte, 2025). However, games are not merely of economic interest, but also a linguistically rich source of information. The significance of English gaming language is being recognized across multiple disciplines. Mašić and Tarabar (2021, p. 33) report that high school students in Bosnia and Herzegovina perceive gaming language as useful, because they can use it for daily face-to-face conversations as well; according to the students, gamers better understand native speakers, find it easier to learn new words and useful phrases via games than in class, and develop their adaptational skills. Sundqvist (2019) observes that Swedish teenage gamers show much greater fluency in the English language than their non-gamer peers. Laato et al. (2021) claim there exists a continuous pressure on the Finnish language to evolve under the influence of English gaming language as speakers increasingly incorporate English

terminology not localized into Finnish. Overall, "the language of computer users is gradually becoming the common language" (Shcherbina et al., 2022, p. 678). These examples illustrate few of the reasons why studying gaming language may prove valuable.

2.2 Knowledge transfer and gaming language as stratified jargon

The stratification of language has an influence on identity formation and personal expressions, because language choices reflect positions and cultural affiliations (Alisoy, 2024). In terms of gaming language, players themselves perceive the stratification within: "it is not surprising that gamers construct their particular social realities about themselves as gamers within their communities of practice in order to ensure mutual understanding and promote a sense of belonging" (Ensslin, 2012, p. 106). Given the volume of games being released, it would not be far-fetched to expect a certain kind of specialization within gaming practice, with gamers focusing on specific game genres. This idea has not yet been academically researched in detail, however, but merely sporadically looked at, mostly in bachelor's and master's theses (e.g., Hsu, 2020; Baňas, 2024). Researchers have studied the individual genres, mostly ones that are commercially successful (Šuster, 2015; Furianto and Simanjuntak, 2023, Badia Climent, 2024, Shahrir Raghbir et al., 2025); however, the relationship between genres remains unexplored.

Naturally, transferability of knowledge within one genre, as well as between genres, does exist. We assume a transferability of knowledge or skills is present in gaming language and gaming overall. This transfer represents specific knowledge or skills that were acquired in one context and can be automatically transferred into a new, similar context (Lieberman et al., 2014). For instance, if a player plays one multiplayer online battle arena (MOBA) game, then it may be expected they will be more effective in other MOBA games, unlike a player who does not have any prior experience with the genre. Such transfer entails shifting pre-existing knowledge and skills into learning, solving new problems or carrying out new objectives within new contexts.

Lieberman et al. (ibid.) distinguish between near transfer and far transfer, and between low-road and high-road transfer. Both dichotomies are relevant here. The near transfer between games that are more alike mechanically is going to be more prevalent and efficient than far transfer between games with lower mechanical similarity – i.e., games of different genres. Low-road transfer "involves the use of well-established skills that are so ingrained that they are easily triggered and occur in an automatic way"; the latter "involves abstracting prior knowledge and applying it to another context, consciously searching for associations between contexts that are hard to recognize" (ibid., p. 3). Use of previously learned specialized vocabulary for purposes of more efficient communication with other players can be understood as knowledge transfer of the low-road type.

Bavelier et al. (2012, p. 392) propose that video games teach "the capability to quickly learn to perform new tasks", the so-called "learning to learn" concept. In Bavelier's view, effective transfer of learning must provide accessible resources, conceptual knowledge and learning rules, which enable players to adapt, generalize and apply what they already know to new situations. Though Bavelier focuses on the experiential-cognitive side of knowledge transfer, a significant part of this learning process takes place via player interaction (and, with an increasing tendency, gamer interaction outside of the game, via so-called meta-game discussions), that is, language-based interaction. Video games' disposition for teaching their players learning to learn should then result in ever-increasing efficiency of game-related communication and specialization of jargons of related games due to higher efficiency of low-road knowledge transfer within speakers.

3. Questions and method

3.1 Hypothesis statement

The aim of this study is to evaluate whether the global Anglophone gaming community uses a single, more or less unified jargon, or whether there exist relatively separate "sub-jargons" that are not necessarily mutually intelligible. So far, theoretical texts on gaming language have treated it an essentially monolithic structure – despite empirical research usually focusing on a single game genre or separating its material on the basis of genres. Showing that a division of gamer speaker communities can be established along the lines of game genres would be significantly informative for future research, from that of gaming language per se, to topics only tangentially related to gaming communication (e.g., second-language acquisition topics, translation and localization studies, natural language processing). Given this study's exploratory nature, we decided to temporarily forego more fine-grain analyses; however, a dataset with an extent of information exceeding the needs of this study was collected in order to accommodate further observations.

We attempt to approach the topic from a lexical-semantic point of view, focusing on the language level most easily reflected on by its speakers. Due to the nature of the knowledge transfer in becoming skilled at games as outlined in Section 2.2, we expect that gamers will be better versed in the vocabulary of the genres they actively play. At the same time, we posit that the process of stratification of gaming language – as a result of stratification of the gaming experience – results in a situation where players will no longer be able to easily comprehend vocabulary outside of their own genres. There will necessarily exist a certain level of inter-genre knowledge transfer, nonetheless, we expect players of a specific genre ($players_G$) to have greater difficulty understanding vocabulary of genres they do not play. A working hypothesis can be outlined as follows:

H: *Gaming language does not function as a single register, but as a cluster of partially independent sociolects. The degree of their mutual intelligibility is primarily determined by the overlap of game mechanics, resulting in high lexical stratification in different genres, rather than a shared universal gamer identity.*

This hypothesis can be tested through an observation of gamers' comprehension of vocabularies tied to different game genres. We decided to use six genres in order to ascertain a meaningful degree of contrast for the purposes of studying jargon stratification, while keeping the number of observed items as low as possible in order to not dissuade potential subjects from participating. Five genres – massively multiplayer online role-playing games (MMORPG), first-person shooters (FPS), multiplayer online battle arenas (MOBA), action-adventure games, and strategy games – were selected based on their cultural prevalence and a relatively small expected terminological overlap, arising from the mechanical differentiation of the play execution within these genres (Enbayeva and Khanzin, 2018, p. 118). We prioritized video game genres that are significant in revenue and market penetration numbers according to available data (Inlingo, 2025, Logrus IT, n.d., Newzoo, n.d.). A sixth – digital tabletop games (DTT) – serves as a control due to its inherently distinct vocabulary, despite its low cultural prevalence, aside from digital collectible card games.

3.2 Lexical-semantic comprehension survey

For the purposes of achieving the research aim, a questionnaire (provided in the supplementary materials) was designed to collect data anonymously from members of the global Anglophone gaming community, and disseminated online, primarily via various *Discord* gaming communities in order to reach the appropriate target audience, but also via other communication platforms, emails and word of mouth. Even though respondents reached through *Discord* are likely to show an above-average level of dedication to the hobby and to not perfectly represent the global population, a sample population with a possible greater degree of gaming vocabulary comprehension suits the needs of this study, particularly for the purposes of verifying our assumptions about individual lexical items (addressed in Section 3.3). While a questionnaire does not allow for directly observing comprehension in communication behaviour, other methods would have their own limitations: corpus-based methods would not permit efficient estimation of comprehension levels, while laboratory-observed tasks would limit sample sizes.

The questionnaire consists of two parts. The first collects traditional demographic data, as well as information on respondents' gaming habits. For the purposes of this paper, with the aim to observe general trends across global gamer communities, these factors are not relevant; these data are intended for analytically estimating which variables may have an impact on comprehension and overall proficiency in future follow-up analyses. However, for illustrative purposes, a description of the sample population based on the data collected in this part is

provided in Section 4.1. The second part represents the vocabulary comprehension survey itself, collecting data on how respondents understand lexical items specific to the selected genres. A structuralist approach to semantic meaning as a set of semantic components (Cruse, 2000, p. 98) was taken, as it allows for an easily categorizable and comparable recording of responses, and thus quantitative analysis of the obtained data. For this purpose, a list of lexical items and their semantic components was established for each of the six selected genres based on both the researchers' personal experience and an extended series of consultations[1] with people heavily vested in individual video game genres. From this list, seven lexical items were selected for use to represent each of the selected genres. For each item, five items – a combination of its necessary semantic components and components not applicable to them were compiled. It is necessary to state that we do not claim an authoritative interpretation of the lexical items' semantic component make-up, nor do they believe it is achievable at all – as Ensslin highlights, "in such fast-moving areas as the videogame industry and games studies, lexical investigations have to be open to rapid changes" (2012, p. 44). Even though the vocabulary list used here is a result of extensive research and consultations with members of the gaming community, it is not possible to delimit exact meanings of the lexical items due to variation in usage. The "answer key" used for this survey is thus understood to hold capacity for deviation from the norm. Due to this, the precision of this set of expected and inapplicable components (further labelled *canonical key*; this key is made available to the readers by providing key ratings within the questionnaire in the supplementary materials) was itself subject to verification in this study by design.

While we expect to notice only random, not systematic overall deviations from the canonical key, there is a possibility that a systematic under- or over-selection of components will take place. Under-selection, or the phenomenon that players$_G$ do not perceive a specific semantic component as a part of meaning complex, is more likely – it could mean that the given component is specific to games the questionnaire creators are familiar with, but not common for the genre-wide use. On the other hand, over-selection of incompatible components is highly unlikely to occur, as it would mean that a word is systematically perceived as having a meaning denotationally opposite to the researchers' and consultants' consensus. Even if the meaning of a word evolves over time, lexical items systematically do not change their meaning by negation (Geeraerts, 1997).

One additional general-vocabulary lexical item was added at the end of this part of the questionnaire, to serve as a sensitivity check in the data collection process. We determined that this item – *a ham sandwich* – should be comprehensible to all respondents regardless of

[1] These consultations were carried out through shared Google spreadsheets and multiple iterations of online discussions. The main focus was centrality/peripherality of various lexical items and necessity/canonicity of individual controversial semantic components.

language proficiency. A failure to determine the appropriateness of its semantic components in line with the canonical key may signal a response significantly increasing noise in the data.

3.2.1 Survey limitations and countermeasures

We anticipate that the scope of the questionnaire may lead to issues with data collection by demotivating respondents from finishing their submissions. Several concessions therefore had to be made in the design process. Notably, the number of lexical items on which respondents are surveyed was kept at a relatively small amount – seven lexical items per genre and a control item, with only five semantic components each are provided for evaluation. This small number results in increased prevalence of noise for the purposes of some statistical measures, notably, confidence intervals for inter-rater agreement estimates may be quite wide given the small number of ratings. Though this is an acceptable limitation in an exploratory study, consistent with current linguistic methodology (cf. e.g., Vasishth and Gelman, 2019; Gries, 2022), it must be noted that individual item-level agreement statistics should be interpreted with caution; more robust inferences are drawn from genre-level analyses.

Despite limiting the scope of the questionnaire, we still anticipate that some respondents may not complete it, either due to losing interest during the process, or due to behavioural patterns generally observable in the case of anonymous online communication. [2] Such responses would create additional noise and potentially skew the results due to the relatively small sample sizes. For this reason, two sensitivity checks were performed. Responses failing to meet at least 80% agreement with the canonical key for the control item (*ham sandwich*) were qualitatively evaluated and potentially marked for omission from the full dataset. Second, a more robust sensitivity check is an omission of respondents systematically under-selecting semantic components. A secondary data set was created by omitting those respondents who selected only a single semantic component for more than two-thirds of lexical items (i.e., at least 5 out of 7) in at least 75% of the genres they reported actively playing. Genres not reported as played were not subject to this procedure, as under-selection in such a scenario can easily be attributed to lack of knowledge of the vocabulary, not an intent on the respondent's part. These criteria were set to relatively conservative values, as a more radical cut-off could reduce the amount of data below a desirable level. However, we are aware that some or all under-selection thus omitted might be a result of other factors (e.g., low English language proficiency, a generally casual approach to games consumption, overestimating one's skill in the genre). As such, all statistical data in Section 4 are reported for both full and reduced data sets.

[2] Researchers' prior experience with online surveys dictated increased caution; future follow-up studies will be executed with a more liberal approach.

3.3 Evaluating the canonical key

A direct and immediate attempt at providing an answer to the question of gaming language's stratification is hindered by the relative opacity and, more notably, instability of the jargon's vocabulary. This is exacerbated by other facts: mechanical specifics of games within a single genre may translate into differences in semantic content of genre-related lexical items; the same lexical form may carry different meanings across genres, etc. Due to these, no questionnaire of this kind can be created with an authoritative answer key. Alongside its primary focus, the research is therefore designed to also serve as a litmus test for the vocabulary/semantic component list itself: instead of treating the list as a definitive, correct key, the intent was for it to be compared against the respondents' answers.

This was done by assessing the reliability of agreement among players_G and reliability of their agreement with the canonical key. This approach allows for evaluating whether those respondents who are relatively likely to have come into contact with the given lexical item perceive its semantic components differently than the researchers' consultants in establishing the canonical key, and if they do so in a systematic manner. For this purpose, Fleiss's kappa (κ_F) was calculated for the relevant portions of the respondent population (Fleiss, 1971, Gwet, 2014).[3] This measure shows whether the level of rater agreement exceeds chance agreement; a significant κ_F suggests that a particular lexical item has an established semantic meaning among the genre's speaker community. More importantly, κ_F was calculated for each player_G sub-population across all genre items to estimate the overall degree of group agreement.

This measure was supplemented by Cohen's kappa (κ_C), which is used to evaluate inter-rater reliability between two raters, to verify whether the players_G agree or disagree systematically with the canonical key (Cohen, 1960, Gwet, 2014). For this purpose, a compound respondent rating was formulated by deriving a group consensus vector for individual lexical items of a given genre and compared against the rating provided by the canonical key. Additionally, McNemar's test (or exact binomial test in the case of low discordance) was carried out both at item level and genre level. Despite this test having a different purpose than κ_C, it serves here as an additional verification of significance for those cases where disagreement is observed on the basis of κ_C, in order to mitigate typical issues with interpreting κ_C. A tendency towards systematic agreement with the key suggests that the key represents the actual communication practice. A systematic disagreement of players_G with the canonical key (showing both a relatively high negative κ_C and significant p for McNemar's test) would more or less clearly suggest that the canonical key does not match the usage reality of the speaker community.

[3] Data collected were treated using LibreOffice Calc 25.2.2, all calculations for this study were carried out using R 4.5.1.

3.4 Evaluating the effect of genre-related experience on vocabulary comprehension

Data obtained for inter-rater agreement evaluations also serve to test the impact of genre experience on vocabulary comprehension. If understanding of the lexical items' meanings is systematic among a speaker community (regardless of it matching the canonical key), there exists a degree of consensus on what these items mean and how they are used in communication, suggesting that the given population does comprehend (and use) the lexis in a particular way. Vice versa, if no systematic agreement can be observed among a sub-population, it can be assumed that the given lexical items are not a part of their vocabulary – which would suggest that these items do not belong to the gaming jargon as a whole. This means that a pairwise comparison of κ_F of players$_G$ and non-players$_G$ can suggest whether these items represent a shared vocabulary for all of them, or just the player$_G$ sub-population. Secondarily, as long as the canonical key is not found to be systematically disagreed with by players$_G$, κ_C can also be calculated for non-players$_G$, and a comparison provides a datapoint for increasing certainty of interpretation of the κ_F comparison. Testable sub-hypotheses can be formulated:

H_{1a}: *For each genre, Fleiss's κ computed among players will be significantly greater than Fleiss's κ among non-players.*

H_{1b}: *For each genre, the canonical key will show significantly higher Cohen's κ with players' responses than with non-players' responses.*

As the main test of whether comprehension of genre-specific vocabulary depends on players' genre involvement, a mixed factorial design and mixed (repeated-measures × between-subjects) analysis of variance was used (cf. Baayen, 2021, p. 242). To facilitate the evaluation, comprehension scores for each item per respondent were calculated as proportions of respondent's ratings agreeing with the canonical key across the five semantic components. Mean comprehension score, computed per respondent for each genre, was used as a dependent variable in the ANOVA design, resulting in six within-subject levels. Whether any given respondent plays the focal genre served as the between-subjects factor. This design tests three effects of substantive interest:

(1) main effect of *genre*, which shows whether some genres' vocabularies are easier or harder to comprehend generally (i.e., across the player community, regardless of experience with the genre);
(2) a main effect of *playing a genre*, which shows whether players$_G$ and non-players$_G$ differ in comprehension overall; and, most importantly for the research design,
(3) the *genre* × *playing genre* interaction, which directly evaluates whether the comprehension profile across genres differs between players and non-players — in

particular, whether players$_G$ perform relatively better on G's vocabulary than non-players.

A third testable sub-hypothesis for this method can be formulated:

H_2: *Mixed ANOVA with genre as 6-level within-subject factor and genre experience as between-subjects factor, there will be a significant genre × genre experience interaction such that players score higher on comprehension of their own genre than non-players.*

ANOVA results are reported with Greenhouse–Geisser correction where Mauchly's test indicates sphericity violation (Verma, 2016, p. 42); effect sizes are reported as partial η^2 with 95% confidence interval.

Together, the mixed ANOVA and inter-rater agreement tests allow to assess both overall comprehension patterns and whether groups of players delimited by their experience with individual genres share distinct, systematic understandings of genre vocabulary. If significant effects are identified and a clear discrepancy in the degree of inter-rater agreement is observed, it would suggest that there is a degree of vocabulary stratification within the gaming jargon that would warrant treatment of gaming language as a set of separate (but not independent) jargons.

4. Results and interpretation

4.1 Description of respondent population

The survey was taken by 252 respondents self-reporting as playing video games. These comprise the primary data set used for calculations, labelled as *full*. The procedure for narrowing down the data set by omission of questionable responses outlined in Section 3.2.1 led to an omission of 27 responses. Calculations carried out using this data set are labelled as *reduced*.[4] More radical thresholds for the procedure would have resulted in larger omissions (33 responses if players$_G$ under-selecting as little as half of their genres were omitted; 37 if the threshold for a genre would be set to 4 instead of 5 under-selected items; 48 responses if both criteria were adjusted as outlined), however, we believe the more conservative thresholds used prevent discrimination of respondents on the basis of lower English language proficiency or relatively lower experience with video games. The overview of the population's basic parameters, habits and behaviours linked to gaming and language is presented in *Table 1*, while an overview of reported played genres is presented in Figures 1 and 2.

[4] Throughout the paper, values calculated from the reduced set are provided in square brackets.

Table 1. Overview of the respondent population

		Full (N = 252)	**Reduced** (N = 225)
gender	male	173	160
	female	62	48
	other	17	17
age	< 18	8	7
	18-24	150	130
	25-34	80	76
	35-44	11	10
	45 +	3	2
native English speaker	yes	54	53
	no	198	173
preferred UI language	English	233	212
	other	19	13
Preferred language for communication with other gamers (in-game, in online platforms)	English	230	209
	other	22	16
Length of gaming experience	< 1 year	4	2
	1-5 years	25	20
	5-10 years	37	31
	> 10 years	186	172
Time spent gaming per week	< 5 hours	50	40
	5-10 hours	72	62
	10-20 hours	63	61
	> 20 hours	67	62
Dominant in-game interaction	Plays alone	59	52
	Plays with IRL friends	102	94
	Plays with online friends	52	47
	Plays with established in-game community	14	12
	Plays with PUGs	25	20
Gaming content consumption	yes	231	210
	no	21	15
Participation in meta-game discussions	Both online and in-person	162	150
	Online only	84	67
	In-person only	10	8
	Does not participate	0	0
Income from gaming-related activities	yes	37	34
	no	215	191

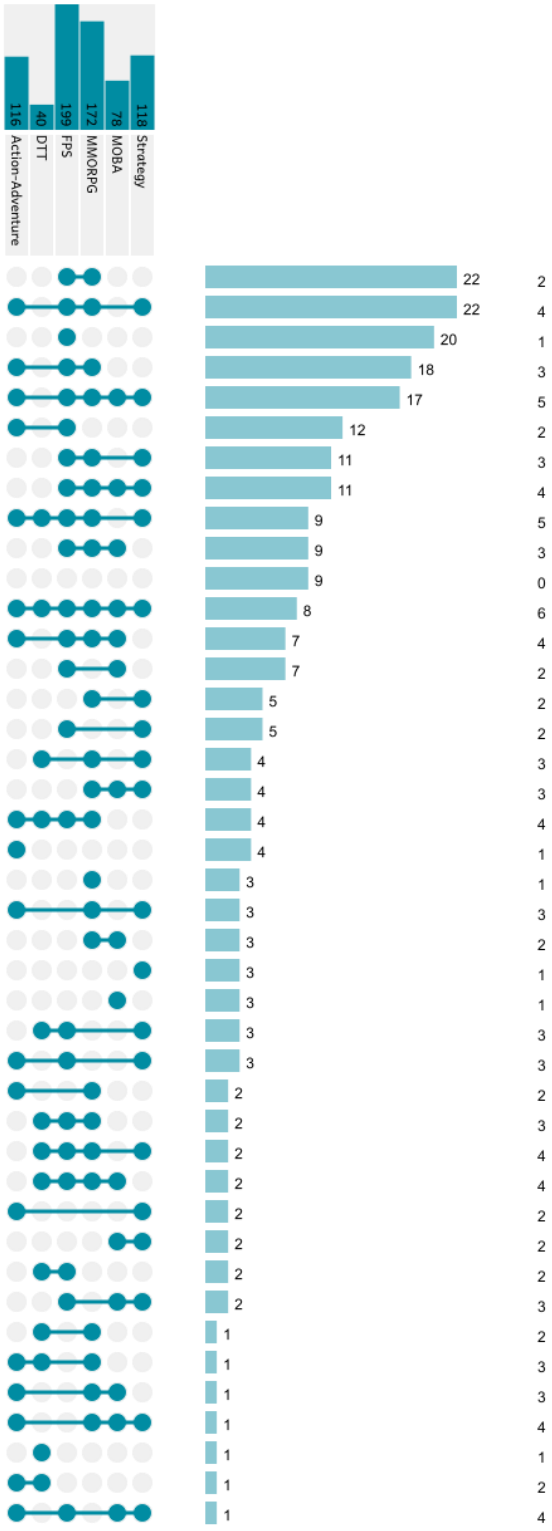

Fig. 1. Reported genres – full data set

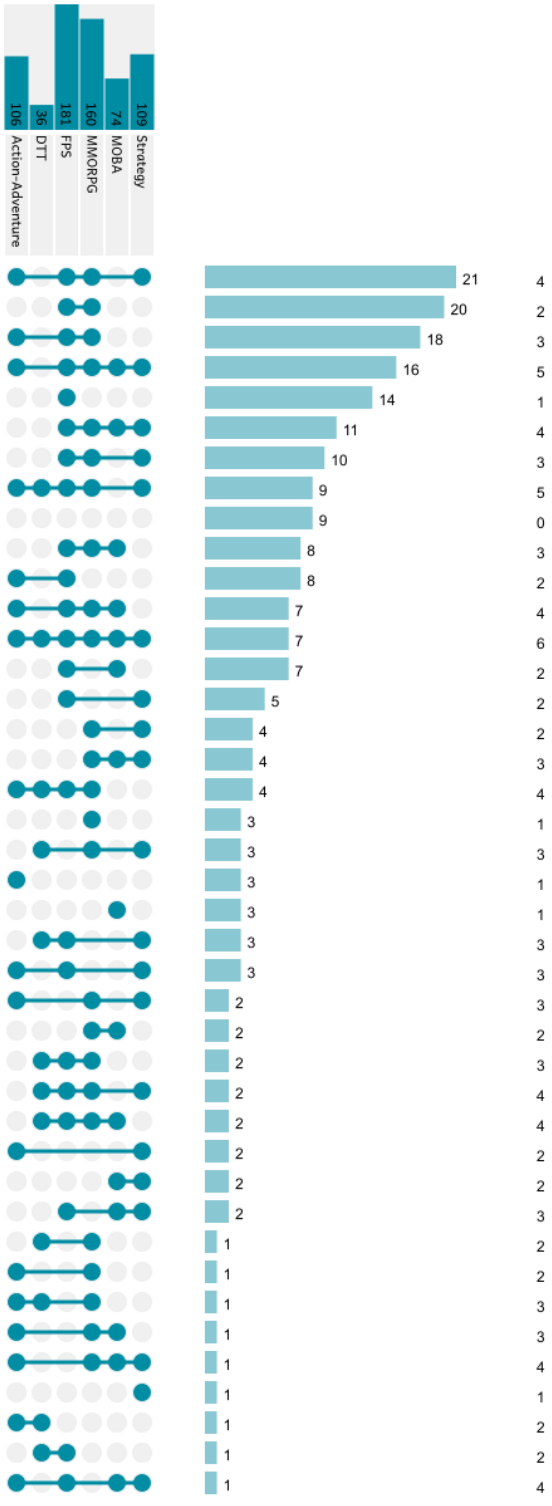

Fig. 2. Reported genres – reduced data set

The information on gaming habits suggests a relatively high average level of respondent dedication to games. Only 19.84% [17.78%] of respondents report playing games for less than 5 hours per week, with 51.58% [54.67%] spending more than 10 hours per week playing (with more than a quarter of respondents dedicating most likely at least 3 hours daily on average to the hobby). At the same time, 73.81% [76.44%] of respondents report having been playing video games for the maximum provided value of at least 10 years (with a significant number of these respondents being no more than 24 years old; and all the older respondents belonging in this group). This is also supported by reported gaming-related habits, both passive content consumption (reported by 91.67% [93.33%] respondents) and active discussion participation, with all respondents participating, and an overwhelming majority – 96.03% [96.44%] doing so regularly over various online platforms.

The data also clearly support the generally accepted assumption of prevalence of English language in gaming discourse. Even though respondents' fluency in English was necessary and at least partial use of English in meta-game communication was predictable due to the mode of survey dissemination, the low preference for respondent's native language in conversation (only 8.73% [7.11%]) and even lower preference for non-English user interface (7.54% [5.78%]) suggest a substantial preference for English regardless of other options.[5]

In omitting respondents under-selecting semantic components, the make-up of population shifted so that the prevalence of high-exposure behaviour rose. Similarly, preference for English language both in-game and in meta-game communication became more pronounced.

Data on genres show that attempts at observing the expected inter-genre knowledge transfer positively impacting comprehension scores will be impeded by the fact that very few gamers deal in one game genre exclusively. Respondents playing more than a half of the observed genres make up 33.33% [36%] of all cases. Calculations for subsets of respondent population meeting genre inclusion/exclusion criteria would require working with very small datasets, and thus very wide confidence intervals. Taking into account also the uncertainty about data credibility resulting from self-reporting, we elected to carry out all calculations on the basis of the single factor of relevant genre experience. The population sizes used in this manner result in κ for each genre being calculated from pooled respondent–item judgments constructed from between 1,260 and 3,000 respondent–item pairs. These large per-genre sample sizes exceed the ≈ 940 paired judgments (at anticipated $\kappa = 0.4$; acceptable width of CI $W = 0.1$, and the prevalence of positives $\pi = 0.5$) recommended by Watson and Petrie (2010, p. 1178) for stable κ estimation. The collapsed consensus vector used as a rater in the calculation is constructed from these underlying respondent-level judgments, meaning each of its 35 entries reflects majority agreement over a large number of raters. Although the

[5] This trend, however, is also explainable by absence of video game localizations, especially for languages with lower numbers of speakers – and future survey iterations should be expanded to include questions relevant to translation studies as well.

consensus–canonical κ_C is therefore formally computed on 35 items, the underlying data supporting each consensus rating are sufficiently extensive.

4.2 Evaluation of the canonical key

In order to evaluate whether the canonical key compiled from consultations deviates from the actual usage, κ_F was calculated for the selected genres using data from player$_G$ sub-population as a preliminary measure of rater consistency within groups. In Table 2, all relevant data used in the evaluation – number of raters N, overall proportion of agreement P, and probability of chance agreement P_e – are reported; estimated κ_F is reported with its lower and upper bounds for 95% confidence interval (CI).

Table 2. κ_F of player$_G$ sub-populations

genre	N	P	P_e	κ_F	95% CI
full					
MMORPG	172	0.684	0.509	0.356	[0.344, 0.368]
FPS	199	0.717	0.508	0.425	[0.414, 0.436]
MOBA	78	0.745	0.503	0.487	[0.471, 0.504]
Action-adventure	116	0.683	0.501	0.364	[0.35, 0.379]
Strategy	118	0.693	0.500	0.387	[0.373, 0.401]
DTT	40	0.630	0.507	0.251	[0.225, 0.276]
reduced					
MMORPG	160	0.697	0.505	0.387	[0.375, 0.399]
FPS	181	0.741	0.504	0.477	[0.466, 0.488]
MOBA	74	0.759	0.502	0.518	[0.501, 0.534]
Action-adventure	106	0.705	0.500	0.410	[0.395, 0.424]
Strategy	109	0.719	0.501	0.437	[0.423, 0.451]
DTT	36	0.647	0.502	0.291	[0.266, 0.316]

On the basis of Landis and Koch (1977), these results show a fair ($\approx$0.21–0.40) to moderate ($\approx$0.41–0.60) degree of agreement; in no category do players$_G$ show a substantial degree of agreement within their group within CI, not even within the reduced data set. This suggests the raters agree among themselves more than by chance, but a certain level of disagreement exists – the interpretation of lexical items' meanings across genres is not highly consistent among populations overall – while players$_G$ share a broadly similar understanding of the evaluated semantic components, substantial variability in understanding individual semantic components remains. The marginal increase in each category caused by the reduction is a

necessary result of eliminating under-selected responses. The results of inter-rater agreement estimates with the canonical key should therefore be treated cautiously.

The actual evaluation of the canonical key was carried out by means of calculating κ_C, treating the canonical key as $rater_1$ and group consensus vectors for $players_G$ as $rater_2$. Contingency tables used for calculations are reported in Table 3 alongside proportionate agreement P_o, probability of chance agreement P_e, κ_C, with lower and upper bounds for 95% CI, and the McNemar's exact binomial test p-value, which we report instead of the more common asymptotic χ^2 approximation due to the total number of discordant pairs being insufficient (< 25) in all comparisons (Fagerland, Lydersen and Laake, 2013).

Table 3. κ_C of $player_G$ sub-populations × canonical key

genre	$r_1=1$; $r_2=1$	$r_1=1$; $r_2=0$	$r_1=0$; $r_2=1$	$r_1=0$; $r_2=0$	P_o	P_e	κ_C	95% CI	p_{exact}
full									
MMORPG	15	7	0	13	0.800	0.482	0.614	[0.378, 0.85]	.016
FPS	14	6	0	15	0.829	0.486	0.667	[0.438, 0.896]	.031
MOBA	18	3	0	14	0.914	0.503	0.828	[0.644, 1]	.250
Action-adventure	15	5	1	14	0.829	0.494	0.661	[0.421, 0.902]	.219
Strategy	21	0	0	14	1.000	0.520	1.000	[1, 1]	N/A
DTT	14	8	2	11	0.714	0.489	0.441	[0.164, 0.718]	.109
reduced									
MMORPG	17	5	0	13	0.857	0.496	0.716	[0.496, 0.937]	.063
FPS	16	4	1	14	0.857	0.498	0.715	[0.488, 0.943]	.375
MOBA	18	3	0	14	0.914	0.503	0.828	[0.644, 1]	.250
Action-adventure	18	2	1	14	0.914	0.506	0.826	[0.639, 1]	1
Strategy	21	0	0	14	1.000	0.520	1.000	[1, 1]	N/A
DTT	15	7	2	11	0.743	0.496	0.489	[0.214, 0.765]	.180

The wide CI margins for κ_C result from the fact that the measure is technically calculated using only 35 paired comparisons – however, both of these are based on consensus of significantly larger groups, as reported earlier. As such, (cautious) interpretations of the estimates are justified.

All sub-populations show a generally good agreement with the canonical key, which is further improved after data omissions. As expected, the lower extreme is the digital tabletop games genre, showing "only" a fair to moderate agreement. It is also worth noting that the

degree of respondent uncertainty remains nearly identical even after data set correction. The noteworthy upper observed extreme is the perfect agreement in case of strategy games. Despite variation in individual ratings evident from κ_F, there is a clear indication that the canonical key captures the meanings recognized by the player community. The remaining four genres all show overall substantial agreement, which is significantly increased in the corrected data set. While the contingency table shows a general asymmetric discordant tendency for under-selecting components, it is also worth noting that the McNemar's test loses its significance for both MMORPG and FPS with data set reduction.

The results presented in Table 3 suggest that players$_G$ generally substantially agree with the canonical key, which can thus be interpreted as a good representation of lexical items as used by the speaker communities. For future use of the lexical items/semantic components list, individual items and their components will be re-evaluated on the basis of component-level P_i (≥ 0.6) and item-level κ_F and κ_C – especially for cases where κ_C is negative, indicating systematic disagreement, regardless of degree. Such a case was observed only for a single item within the DTT lexicon: *dice roll modifier*, for which two components, [*choice*] and [*randomness-enhancing*] show noteworthy discordant rating. An immediate adjustment of the key was considered for this item; however, we decided to retain the key unmodified for purposes of this paper: this case may simply result from the expected lower comprehension of the DTT lexicon given its systematic mechanical difference from other sub-jargons. Additionally, these components are provably a canonical and non-applicable part of its semantic meaning complex, respectively, as understood within the wider tabletop gaming discourse and illustratable on specific board game designs.

4.3 Comprehension variation between players and non-players

With the fit of canonical key verified, κ_F and κ_C were calculated also for non-players$_G$ and compared with player$_G$ results. Results for these measures are presented in Tables 4 and 5.

Non-players show a less stable comprehension of genre vocabularies than players. While this observation is predictable, the degree of difference is rather sizeable – most pairs of κ suggest a one-step lower degree of agreement. The smallest discrepancy between κ_F in players and non-players for both full and reduced data sets is seen in the action-adventure genre, which suggests this genre has the lowest degree of lexical specialization from among the observed. Vice versa, the greatest discrepancy is observable in MOBA, suggesting a more profound specialization of vocabulary. These results partially support sub-hypothesis H_{1a}, which appears to apply to 5 out of 6 observed genres.

Table 4. κ_F of non-player$_G$ sub-populations

genre	N	P	P_e	κ_F	95% CI
full					
MMORPG	91	0.650	0.572	0.181	[0.164, 0.197]
FPS	64	0.664	0.566	0.226	[0.207, 0.246]
MOBA	185	0.651	0.547	0.229	[0.217, 0.241]
Action-adventure	147	0.672	0.525	0.309	[0.296, 0.322]
Strategy	145	0.641	0.524	0.245	[0.232, 0.258]
DTT	223	0.602	0.539	0.136	[0.125, 0.146]
reduced					
MMORPG	65	0.666	0.529	0.291	[0.271, 0.310]
FPS	44	0.687	0.515	0.355	[0.345, 0.366]
MOBA	151	0.663	0.524	0.292	[0.276, 0.308]
Action-adventure	119	0.697	0.506	0.387	[0.373, 0.401]
Strategy	116	0.661	0.505	0.316	[0.301, 0.330]
DTT	189	0.605	0.520	0.178	[0.152, 0.203]

Table 5. κ_C of non-player$_G$ sub-populations × canonical key

genre	$r_1=1$; $r_2=1$	$r_1=1$; $r_2=0$	$r_1=0$; $r_2=1$	$r_1=0$; $r_2=0$	P_o	P_e	κ_C	95% CI	p_{exact}
full									
MMORPG	11	11	0	13	0.686	0.452	0.426	[0.196, 0.656]	.001
FPS	12	8	0	15	0.771	0.478	0.563	[0.323, 0.802]	.008
MOBA	10	11	0	14	0.686	0.457	0.421	[0.19, 0.652]	.001
Action-adventure	15	5	1	14	0.829	0.494	0.661	[0.421, 0.902]	.219
Strategy	18	3	0	14	0.914	0.503	0.828	[0.644, 1]	.250
DTT	9	13	1	12	0.600	0.445	0.279	[0.045, 0.514]	.002
reduced									
MMORPG	12	0	10	13	0.714	0.460	0.471	[0.236, 0.706]	.002
FPS	13	0	7	15	0.800	0.482	0.614	[0.378, 0.85]	.016
MOBA	12	0	9	14	0.743	0.469	0.516	[0.278, 0.755]	.004
Action-adventure	15	1	5	14	0.829	0.494	0.661	[0.421, 0.902]	.219
Strategy	20	0	1	14	0.971	0.514	0.941	[0.828, 1]	1
DTT	10	1	12	12	0.629	0.452	0.322	[0.079, 0.565]	.003

These observations are also supported by the differences in κ_C values. Agreement with canonical key is the same for $players_{action\text{-}adventure}$ and $non\text{-}players_{action\text{-}adventure}$, while $non\text{-}players_{MOBA}$ show the greatest degree of reduction – and a markedly lower agreement with the canonical key than $players_{MOBA}$. The difference between groups is slightly less prominent for FPS, but still signals a reduction by one level in the traditional interpretation scale. These results therefore partially support sub-hypothesis H_{1b}, with a discernible pattern for action-adventure genre deviating from the trend observed for other genres.

This information may be correlated with interpretation of results of the mixed analyses of variance calculated for each of the six selected genres. These are presented summarily in Table 6. Values of partial η^2 are interpreted for effect sizes on the basis of Cohen's (1988) widely adopted thresholds (Lakens, 2013; Richardson, 2011).

Table 6 – mixed ANOVA results

	F	df	p	ηp2	95% CI
full					
MMORPG					
Player $Status_{MMORPG}$	30.5	1	<.001	0.109	[0.047, 0.185]
Genre†	109.59	4.20	<.001	0.305	[0.264, 0.342]
Player $Status_{MMORPG}$ × Genre†	2.21	4.20	.062	0.009	[0.000, 0.018]
FPS					
Player $Status_{FPS}$	11.8	1	<.001	0.045	[0.008, 0.105]
Genre†	81.65	4.27	<.001	0.246	[0.206, 0.283]
Player $Status_{FPS}$ × Genre†	5.07	4.27	<.001	0.020	[0.005, 0.034]
MOBA					
Player $Status_{MOBA}$	11.4	1	<.001	0.043	[0.008, 0.103]
Genre†	120.6	4.44	<.001	0.325	[0.284, 0.362]
Player $Status_{MOBA}$ × Genre†	18.7	4.44	<.001	0.070	[0.042, 0.095]
Action-Adventure					
Player $Status_{Action\text{-}Adventure}$	4.49	1	.035	0.018	[0.000, 0.063]
Genre†	120.591	4.18	<.001	0.325	[0.284, 0.362]
Player $Status_{Action\text{-}Adventure}$ × Genre†	0.344	4.18	.857	0.001	[0.000, 0.003]
Strategy					
Player $Status_{Strategy}$	16.6	1	<.001	0.062	[0.017, 0.128]
Genre†	123.43	4.18	<.001	0.331	[0.289, 0.368]
Player $Status_{Strategy}$ × Genre†	4.2	4.18	.002	0.017	[0.003, 0.030]
DTT					
Player $Status_{DTT}$	6.88	1	.009	0.027	[0.002, 0.078]
Genre†	61.004	4.18	<.001	0.196	[0.157, 0.232]
Player $Status_{DTT}$ × Genre†	0.402	4.18	.816	0.002	[0.000, 0.004]

reduced					
MMORPG					
Player Status$_{MMORPG}$	21.3	1	<.001	0.087	[0.029, 0.164]
Genre†	107	4.12	<.001	0.324	[0.281, 0.363]
Player Status$_{MMORPG}$ × Genre†	2.85	4.12	.022	0.013	[0.000, 0.025]
FPS					
Player Status$_{FPS}$	8.68	1	.004	0.037	[0.004, 0.098]
Genre†	78.48	4.20	<.001	0.260	[0.217, 0.300]
Player Status$_{FPS}$ × Genre†	5.21	4.20	<.001	0.023	[0.006, 0.039]
MOBA					
Player Status$_{MOBA}$	6.5	1	.011	0.028	[0.001, 0.084]
Genre†	120.5	4.41	<.001	0.351	[0.307, 0.390]
Player Status$_{MOBA}$ × Genre†	18.5	4.41	<.001	0.077	[0.046, 0.105]
Action-Adventure					
Player Status$_{Action\text{-}Adventure}$	3.07	1	.081	0.014	[0.000, 0.058]
Genre†	119.389	4.11	<.001	0.349	[0.305, 0.387]
Player Status$_{Action\text{-}Adventure}$ × Genre†	0.302	4.11	.881	0.001	[0.000, 0.003]
Strategy					
Player Status$_{Strategy}$	16.0	1	<.001	0.067	[0.018, 0.139]
Genre†	122.22	4.10	<.001	0.354	[0.311, 0.393]
Player Status$_{Strategy}$ × Genre†	4.4	4.10	.001	0.019	[0.004, 0.034]
DTT					
Player Status$_{DTT}$	9.36	1	.002	0.040	[0.005, 0.102]
Genre†	59.469	4.11	<.001	0.211	[0.168, 0.249]
Player Status$_{DTT}$ × Genre†	0.707	4.11	.591	0.003	[0.000, 0.008]

† Data reported with Greenhouse-Geisser correction.

Three findings emerge from the mixed ANOVAs. First, genre was highly significant (all $p < .001$) with large partial η^2 values, indicating systematic differences in comprehension across genres. Second, between-subjects effects *predict higher mean comprehension* in player population: players$_G$ scored higher on average than non-players$_G$ in most cases, indicating small-to-medium overall advantages for players$_G$. In case of action-adventure and DTT, p-values are slightly higher, though still significant, and effect sizes indicate only a small advantage, if any (CI for between-group difference for action-adventure includes 0). Notably, between-subject effect is no longer significant for action-adventure in the reduced data set. This is an interesting finding, as the reduction was expected to reduce noise and, therefore, most likely increase the significance of results. In the context of other partial results tied to this genre, it could be interpreted as the comprehension level of action-adventure vocabulary not being tied to experience with the genre.

Third, consistency of findings across genres decreases for interaction terms (player status × genre). For the full data set, interactions were significant for FPS, MOBA, and strategy,

with small-to medium effect sizes. No significance was observed for MMORPG, action-adventure, and DTT. Reduced data set indicates significance also for MMORPG, though with a very small effect size (CI includes 0). Effect sizes for other significant interaction terms increased marginally with data set reduction. This indicates that the genre profile of comprehension differs between players and non-players for some, but not all genres. Evidence for sub-hypothesis H_2 is therefore limited, though not insignificant. The small effect sizes for factor interactions could potentially be explained as resulting from the respondents' high degree of dedication to the hobby in general, as reported in Section 4.1. It is likely that a larger sample size with a more balanced demographic make-up would show greater effect sizes.

4.4 Summary of findings

All three sub-hypotheses were partially supported, providing limited evidence for the research hypothesis. Out of the six observed genres – MMORPG, FPS, MOBA and strategy – findings are consistent with hypotheses, showing that genre produces large differences in comprehension, and players show higher mean comprehension than non-players. For two genres – action-adventure and DTT – no conclusive evidence supporting the main hypothesis was found. This was partially expected in the case of DTT, which was included in the research to see if its high specificity and unique vocabulary will result in lower comprehension overall. We therefore conclude that a certain degree of gaming language stratification into individual, partially distinct jargons does take place for some genres, but this tendency does not seem to be uniform for all genres, and some show equal comprehension rates for the entire speaker community. However, effect sizes for interactions were generally small, so conclusions should be cautious.

5. Conclusion

This pilot study aimed to obtain lexicological evidence to show that gaming language, which is traditionally treated as a single register, consists of multiple, partially independent jargons. We used a survey method to obtain quantitative data on gaming community's comprehension of lexical items connected to six different game genres. Generally, it was observed that genres have a significant effect on comprehension, with players showing higher mean comprehension than non-players. Obtained results indicate that there exist varying degrees of stratification to lexicons of individual game genres. Findings suggest that for four observed genres – MMORPG, FPS, MOBA and strategy games, partially separate lexical subsystems exist. The results are not universal, though, and two of the observed genres – action-adventure and digital tabletop games were not found to differ in comprehension by players and non-players. These results have implications not only for further lexicological and lexicographic research on gaming language, but also for other fields dealing with gaming discourse. The study is limited by 1) limited item-level depth of data due to the deliberately wide, exploratory sampling, and 2)

demographically unbalanced respondent population, skewed towards high degree of player experience and community participation. As a pilot investigation, the study provides preliminary observations about genre-based stratification that can be used in further research into systemic relations between game genre jargons.

Acknowledgments

This research was supported by the Slovak Research and Development Agency, Slovak Republic grant APVV 23-0223.

References

Alisoy, H. (2024). *Stratification of language and society*. n. p. https://doi.org/10.5281/zenodo.10507793

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge University Press.

Badia Climent, S. (2024). *Los MOBA vs las conversaciones cotidianas: análisis de la dinámica interactiva y su extensión estructural. Pragmalingüística*, (32), 71–98. https://doi.org/10.25267/Pragmalinguistica.2024.i32.03

Baňas, J. (2024). *The relationship between active gaming and the understanding of English expressions originating from videogames* [diploma thesis]. University of Prešov. 68 p.

Bavelier, D., Green, C., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience 35*. 391–416. https:doi.org/10.1146/annurev-neuro-060909-152832

Bussmann, H. (1996). *Routledge dictionary of language and linguistics*. Routledge.

Cerezo-Pizarro, M., Revuelta-Domínguez, F.-I., Guerra-Antequera, J., & Melo-Sánchez, J. (2023). The cultural impact of video games: A systematic review of the literature. *Education Sciences, 13*(11), 1116. https://doi.org/10.3390/educsci13111116

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum.

Cruse, D. A. (2000). *Meaning in language: An introduction to semantics and pragmatics*. OUP.

Duarte, F. (2025, March 28). *How many gamers are there? (New 2025 statistics)*. https://explodingtopics.com/blog/number-of-gamers

Enbayeva, L. V., & Khanzin, Yu. S. (2018). *Discoursive Charac-Teristics of Computer Games: A Comparative Analysis*. In PNRPU Linguistics and Pedagogy Bulletin. https://doi.org/10.15593/2224-9389/2018.4.10

Ensslin, A. (2012). *The language of gaming*. Palgrave Macmillan.

Fagerland, M. W., Lydersen, S., & Laake, P. (2013). The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC Med Res Methodol 13*(91). https://doi.org/10.1186/1471-2288-13-91

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378–382. https://doi.org/10.1037/h0031619

Furianto and Simanjuntak, R. R. (2023*). Gaming Language as Language Variations in Digital Humanities*. E3S Web of Conferences 388, 04010 (2023). https://doi.org/10.1051/e3sconf/202338804010

Gries, S. (2022). *Toward more careful corpus statistics: uncertainty estimates for frequencies, dispersions, association measures, and more*. In Research Methods in Applied Linguistics. https://doi.org/10.1016/j.rmal.2021.100002

Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics.

Hsu, C. (2020). *Gaming Language and How It Is Characterized*. Lund University. 72 p. https://www.lunduniversity.lu.se/lup/publication/9031632

Inlingo. (2025, January 10). *The Most Popular Game Genres: What Gamers and Developers Are Choosing in 2025*. https://inlingogames.com/blog/the-most-popular-game-genres/

Laato, S., Inaba, N., Paloheimo, M., & Laajala, T. (2021). Group polarisation among location-based game players: An analysis of use and attitudes towards game slang. *Internet Research, 31*(5), 1695–1717. https://doi.org/10.1108/INTR-03-2020-0158

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.

Landis, J. R., & Koch, G. G. (1977), The measurement of observer agreement for categorical data. *Biometrics 33*(1), 159–174. https://doi.org/10.2307/2529310

Lieberman, D., Biely, E., Thai, C., & Peinado, S. (2014). Transfer of learning from video game play to the classroom. In Blumberg, C. (ed.), *Learning by playing: Video gaming in Education* (pp.189–204). OUP. https://doi.org/10.1093/acprof:osobl/9780199896646.003.0013

Logrus IT. (2024). *Video game market trends in 2024*. https://games.logrusit.com/en/news/game-industry-trends-2024/

Mašić, A., & Tarabar, A. (2021). The influence of online games on learning English vocabulary in high schools in Bosnia and Herzegovina. *MAP Education and Humanities, 1*. 28–37. https:doi.org/10.53880/2744-2373.2021.1.1.28

Newzoo. (n.d.). *Global Games Market Report 2025: Free version*. https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2025

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147.

Shahrir Raghbir, R. S. A., Ng, L.-L., & Chew, S. Y. (2025). *Relating behavioural intentions towards collaborative discussions and language performance of ESL online game players in MMORPG context.* 3L: Language, Linguistics, Literature, 31(3), 161–183. https://doi.org/10.17576/3L-2025-3103-10

Shcherbina, V. E., Pasechnaya, L. A., Simutova, O. P., Verzhinskaya, I. V., & Belova, N. A. (2022). The lexicon of the gaming community members as a youth slang component. In *Social and Cultural Transformations in the Context of Modern Globalism (SCTCMG 2022), Proceedings of the International Conference «Social and Cultural Transformations in the Context of Modern Globalism» (SCTCMG 2022), 19-21 April, 2022, Grozny, Chechen Republic, Russia.* 678–686. https://doi.org/10.15405/epsbs.2022.11.92

Shliakhovchuk, E. (2024). Video games as awareness raisers, attitude changers, and agents of social change. *International Journal of Computer Games Technology*, 3274715. https://doi.org/10.1155/2024/3274715

Statista (2025, October). *Video games – worldwide market size*. https://www.statista.com/outlook/amo/media/games/worldwide

Sundqvist, P. (2019). *Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary*. *Language Learning and Technology, 23*. 87–113. https://doi.org/10.64152/10125/44674

Šuster, S. (2015). *An investigation into language complexity of World-of-Warcraft game-external texts*. https://doi.org/10.48550/arXiv.1502.02655

Vasishth, S., & Gelman, A. (2019). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. In Linguistics (Vol. 59, pp. 1311–1342). https://doi.org/10.31234/osf.io/zcf8s

Verma, J. P. (2016). *Repeated measures design for empirical researchers*. Wiley.

Watson, P. F., & Petrie, A. (2010). Method agreement analysis: A review of correct methodology. *Theriogenology 73*(9), 1167–1179. https://doi.org/10.1016/j.theriogenology.2010.01.003