Identifying Linguistically Relevant Communities of Practice on Twitch

Heather Burnett
Laboratoire de Linguistique Formelle (UMR 7110)
CNRS - Université Paris Cité, France
heather.burnett@cnrs.fr

Abstract

This paper argues that, when it comes to modeling language variation and change on the video game streaming platform Twitch, it is necessary to consider "meso-level" communities of practice, i.e. communities of practice that are smaller than the full video game community, yet larger than the usual level of analysis in recent linguistics studies: communities associated with individual Twitch channels. We present a computational method for identifying these linguistically relevant communities of practice and show how this method can be useful for analyzing quantitative patterns of sociolinguistic variation in a corpus composed of the chat transcripts of 15 streamers of the game *Elden Ring: Nightreign*.

Keywords: communities of practice, video game language, Twitch, clustering, corpus linguistics

1. Introduction

This paper presents a method for identifying linguistically relevant *communities of practice* (CofP) online, with a specific focus on the video game streaming platform *Twitch.tv*. By linguistically relevant CofPs, we mean the kind of social structures that are appropriate for understanding how language varies and changes according to the social context.

How the use of socially meaningful linguistic form varies, and how new forms emerge, are major research questions in sociolinguistics. From the early days of quantitative (variationist) sociolinguistics (see Labov (1973)), researchers observed that the use of sociolinguistic variables (grammatical alternations, for example, pronouncing the final consonant in the word *working* as [ŋ] vs [n]) are conditioned by aspects of speakers' places in their speech community. In the "first wave" of studies of linguistic variation, the "speech community" was taken to be a whole city (New York (Labov, 1966), Montréal (Sankoff and Cedergren, 1972), Detroit (Wolfram, 1969) etc.) and a speaker's place in this community was established through their demographic properties: usually age, gender, social class and race (see Eckert (2012) for a description of the different "waves" of quantitative sociolinguistic study). However, later more anthropologically oriented studies showed that many sociolinguistic variables do not break down according to demographic properties, and their use is better analyzed as resulting from speakers' places in smaller, locally defined social structures: communities of practice. Introduced into sociolinguistics from education research by Eckert and McConnell-Ginet (1992), a community of practice is "is an aggregate of people who come together around mutual engagement in an endeavor. [...] As a social construct, a community of practice is different from the traditional community, primarily because it is defined simultaneously by its membership and by the practice in which that membership engages" (Eckert and McConnell-Ginet, 1992, 464). Eckert (2000)'s foundational work on language variation as a social practice shows how CofPs can be helpful for understanding the distribution of both phonological sociolinguistic variables (vowel pronunciations) and syntactic ones (negation and negative concord) in a Detroit High School.

In the past 30 years, CofPs have become one of the main units of analysis within both quantitative and qualitative sociolinguistics, studying language a wide variety of social settings. In an influential review of the concept, Meyerhoff (2013), building on Wenger (1998), describes three properties defining CofPs. The first is mutual engagement; that is, "the members of a CofP need to get together in order to engage in their shared practices" (Meyerhoff, 2013, 527). The second is that members share some jointly negotiated enterprise: "members get together for some purpose and this purpose is defined through their pursuit of it " (Meyerhoff, 2013, 528). The final defining criterion of a CofP is that members have a shared repertoire, and "these resources (linguistic or otherwise) are the cumulative result of internal negotiations" (Meyerhoff, 2013, 528). It is through this notion of "negotiated shared repertoire" that CofPs have been found to be especially useful for studying certain kinds of language variation and change in the offline world (see, for example, Bucholtz (1999); Holmes et al. (1999), among others).

Communities of practice have also been identi-

Proceedings of the 1st SoCon Workshop & the 2nd NLPSI Workshop @ LREC 2026, pages 25–33
12 May 2026.

fied as relevant for studying language online. Already a decade ago, Angouri (2015) summarized research on digital communication using CofPs; however, one of the most detailed discussions of whether/how this concept can be fruitfully applied to language variation and change in an online setting is Leuckert and Leuckert (2020). In this paper, which focuses on the news and social media platform Reddit, the authors argue that, although there are some differences with offline CofPs, communities associated with individual Reddit forums, known as *subreddits*, generally meet the criteria laid out above for being considered communities of practice. Through an in depth study of three subreddits: R/LEAGUEOFLEGENDS (a forum devoted to the video game *League of Legends* (Riot, 2009)); R/LINGUISTICS (an academic forum devoted talking about language); and R/RUPAULSDRAGRACE (a forum devoted to discussing the TV show *RuPaul's Drag Race* (LogoTV, 2019), Leuckert & Leuckert argue that participants in these subreddits display mutual engagement, participate in a jointly negotiated enterprise and have a shared repertoire. For example, users on the R/RUPAULSDRAGRACE forum use linguistic expressions like *RuSponsibly* and *ruveal*, which are not found in the other subreddits studied.

Leuckert and Leuckert (2020)'s claim that individual subreddits can be communities of practice is both intuitive and well argued. It is also mirrored in other work that studies language (and other kinds of behavior) on a platform with a similar structure: *Twitch.tv*. Twitch.tv is a video streaming platform, which pairs live video of an individual broadcasting an activity (called the *streamer*), usually playing a video game, with a synchronous live chat, where audience members (*viewers*) can send public messages to be consumed by both the streamer and the other viewers. Similar to the way in which Reddit is composed of subreddits/forums, Twitch is made up of channels associated with individual streamers. Given that these channels and their viewers are named "communities", and that they are often very clearly sites of mutual engagement (learning more about a video game, fostering a sense of community, etc. (Hamilton et al., 2014; Gros et al., 2018)), analyzing Twitch channels as CofPs is a common when it comes to studying linguistic behavior on the platform, see Graham (2019); van der Aa (2021) among others.

Our goal in this paper is to argue that, while individual Twitch channels undeniably show some of the properties of CofPs, when it comes to modeling language variation and change on the platform, the "micro" level, i.e. channels, is sometimes too fine-grained to properly capture the **shared repertoires** that members of online communities have. To illustrate this point, we present a new quantitative study of lexical variation in a corpus composed of chat transcripts from 15 streamers of the video game *Elden Ring: Nightreign* (2025, FromSoftware) on Twitch. We show that accurately capturing the distributions of these sociolinguistic variables in the corpus requires a level of analysis that is intermediary between the whole "Elden Ring" community (exemplified by all 15 streamers) and the chats of each individual streamer. Building on recent work applying similarity and cluster analyses to linguistic data, we provide a method for identifying "meso-level" communities of practice that, we show, allow for a better understanding of how new linguistic forms emerge, are used, and change in gaming communities.

2. The Nightreign Corpus

When writing an article studying gaming communities (on Twitch or elsewhere) it is common to start by speaking of a "macro-level" community: *The [Name of Game] Community* (see, for example, Cochran et al. (2023) for *The League of Legends Community*; Marlatt (2020) for *The Fortnite Community*, etc.). This paper will be no exception: its focus will be on the community often described as the *Souls* community, the *Soulsborne* community or the *Elden Ring Community*. These definite descriptions are often used by players who are fans of the series of games developed by the Japanese studio FromSoftware (*Demon's Souls, Dark Souls 1, Dark Souls 2, Dark Souls 3, Bloodborne, Sekiro, Elden Ring, Elden Ring: Nightreign*) and games with similar structures. The fact that players often use the singular definite description to refer to the Soulsborne community (or similar) suggests that they imagine themselves as being part of a single unified group[1] (Anderson, 1983), but, in the details, it's not so clear who exactly belongs to this "community". For example, 2022's Elden Ring has sold over 30 million copies[2], but it is unlikely that anyone who uses the term "the Elden Ring community" would consider it as having more than 30 million members. To operationalize this macro-level notion of *The Elden Ring Community* in a corpus study, we therefore decided to anchor our study around a particular group of streamers: those who took part in the fourth season of Elden Ring Bingo in 2024 (https://bingobrawlers.com/). Elden Ring Bingo is a competitive game that pits two players against each other to see how fast they can complete a series of objectives in the main single player Elden Ring game (completing lines on a

[1] See for example https://www.reddit.com/r/Eldenring/comments/pa3iwn/do_you_think_the_soulsborne_community_is_toxic/

[2] https://www.thegamer.com/elden-ring-sells-30-million-copies-switch-2/

Streamer	Followers	Nationality
LilAggy	262k	USA
Ginomachino	162k	Canada
Parkenharbor	144k	Canada
Bushy	98k	USA
Captain_domo	47k	USA/Germany
ItzCBD	47k	USA
star0chris	41k	USA
Vswed	34k	Canada
Adef	32k	USA
Mitchriz	31k	USA
Blanxz	18k	USA
Zoodle	18k	Germany
nuclearpastatom	14k	USA
yojosherino	14k	USA
DrDoot	7k	USA

Table 1: The 15 streamers featured in the Nightreign corpus (follower count from July 2025)

bingo board). Additionally, most of these streamers perform Elden Ring themed events at the very popular GDQ (*Games Done Quick*) speedrunning events (Blanxz 2024, Mitchriz 2023, 2024, 2025, Captain_domo 2024, 2025, LilAggy 2024, 2025, star0chris 2025, yojosherino 2025, adef 2025, DrDoot 2025). This makes them identifiable as representatives of the (Western) *Elden Ring Community* to players of other games.

Our corpus is composed of the chat transcripts from the streamers who both participated in the tournament and streamed the newest FromSoftware game, *Elden Ring: Nightreign*, in summer 2025 (see Table 1). All streamers stream in English and enforce an "English only" policy in their chats, regardless of their native language.

More specifically, the *Elden Ring: Nightreign Twitch Chat* corpus is composed of transcripts from all the chats during the period in which *Elden Ring: Nightreign* was actively streamed on Twitch: May 30 (launch of the game) - August 31 2025. The individual chat files were downloaded as .csv files using the *Twitch Chat Downloader* application (https://www.twitchchatdownloader.com/).
In all, the corpus is composed of the chats from 448 streams, and contains productions from 50 130 users (henceforth known as *chatters*). The Elden Ring: Nightreign Twitch Chat corpus is available at https://osf.io/ct3zr/overview?view_only=395154716b8c4d5c9cbb955755a99b8c.

3. Ludolexical variation

One of the best-known features of gaming communities, which is also one of the defining characteristics of communities of practice, is that they often give rise to unique shared linguistic norms. For gamers, this usually involves having a specific vocabulary for game-play (ludic) elements in the game (Iorio, 2010; Ensslin, 2012; Zähres, 2019, among many others). The "Elden Ring Community" is no exception, with community invented terms being widely used, especially when it comes to referring to enemies in the game. For example, consider the message in (1), sent by a user in Zoodle's chat.

(1) I love Elden Ring lingo so much, we got nights cav, BBH, BBK, DTS, RadaBeast, Gargs and so much more (chatter in Zoodle stream 2498943069)

As this chatter illustrates, "Elden Ring lingo" involves linguistic practices such as referring to the enemy *Night's Cavalry* as *nights cav* (employing a morphological process known as clipping), *Bell Bearing Hunter* as *BBH* (acronymization), *Black-Blade Kindred* as *BBK*, *Draconic Tree Sentinel* as *DTS* (also acronymization), *Valiant Gargoyles* as *gargs* (clipping) etc.

In this short paper, we present two examples of variation in "Elden Ring lingo" which, we argue, illustrate our main point: some linguistically relevant communities of practice (at least on Twitch) lie between the macro-level, the *Elden Ring Community*, in this case instantiated by all 15 streamers, and the micro-level, i.e. communities around individual streamers' channels. More specifically, we will study variation in the naming of the most recent Elden Ring game (*Nightreign*) and references to one of the main antagonists in the game: the *Bell Bearing Hunter*.

3.1. Nightreign

We extracted all references to the game *Nightreign* from the *Elden Ring: Nightreign Twitch Chat* corpus. Through searching with the regular expression *night+*, we extracted all the uses (with various spellings) of the official name of the game: *Nightreign* (2). From inspecting the contexts surrounding the occurrences of *Nightreign* (chat messages shortly before and after the original message), we were able to identify two other variants: the acronyms *NR* (3) and *NTR* (4).

(2) Hitless **nightreign** runs are about to be crazy (chatter in Blanxz stream 2471884069)

(3) Chat did **NR** get a new patch today?? (chatter in LilAggy's stream 2482675698)

(4) aggy is the best **NTR** bingo player for sure (chatter in LilAggy's stream 2538401141)

Table 2 shows the distribution of these three variants in the whole corpus. The official name of the game is the most frequent variant, appearing in almost 92% of the cases. However, there is variation:

the acronym *NR* is used around 7% of the time, and the minority variant, *NTR*, is used less than 1% of the time.

Variant	Occurrences
Nightreign	7112 (91.9%)
NR	567 (7.3%)
NTR	59 (0.7%)

Table 2: Variation referring to *Elden Ring: Nightreign*

The appearance of acronyms in a corpus of chat between gamers is not surprising: research on language and video games has shown repeatedly that shortening morphological processes, especially acronymization, is very common (Ensslin, 2012; Zähres, 2019; Jannah and Tawami, 2024, among many others). However, what is less expected, given previous literature, is the existence of two acronym variants: *NR* and *NTR*. A natural hypothesis to explain this state of affairs is that *NTR* is favored by members of a particular streamer's community, as a way of marking the "in group" status, something that has also been argued to be frequent in gaming communities (Ensslin, 2012; Zähres, 2019). However, if we break the data down by individual streamers' chats, we see that this is not so clearly the case. Figure 1 shows the raw counts of each variant and Figure 2 shows the percentage of the total distribution of the variant, for the chats of each of the 15 streamers.

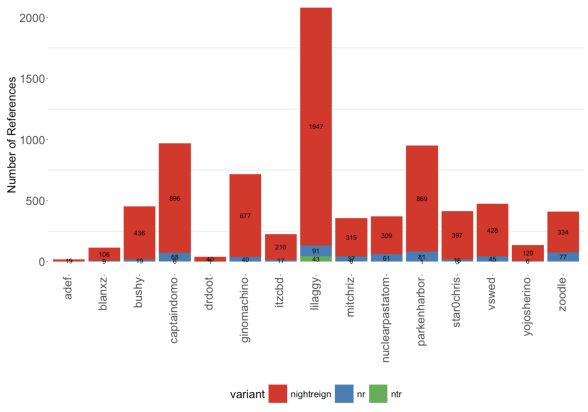

Figure 1: Nightreign vs NR vs NTR by streamer (counts)

This figure shows that *NTR* is used by chatters who watch a number of streamers: Captain_domo, LilAggy, Mitchriz, parkenharbor and star0chris. So a question arises as to whether these five streamers' chats have something in common that could explain the emergence of the *NTR* variant.

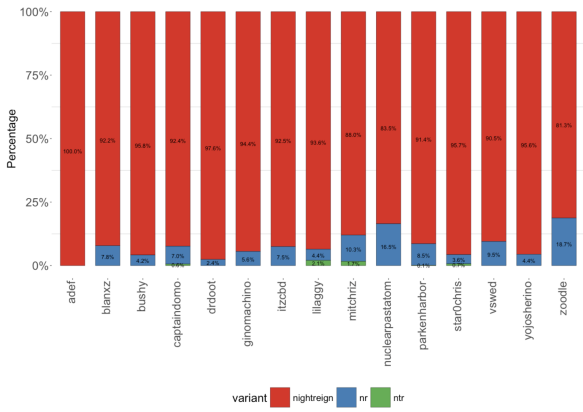

Figure 2: Nightreign vs NR vs NTR by streamer (percentages)

3.2. Bell Bearing Hunter

Our second example is reference to the enemy officially known as the *Bell Bearing Hunter*. Using regular expressions, we extracted all occurrences of use of the variant *bell bearing hunter* from the corpus (5), as well as elliptical versions, such as *bell bearing* (6) and *bell hunter* (7). Inspecting the context around occurrences of *bell bearing hunter*, we further identified other variants[3]: the acronym *BBH* (8), see also (1), and the expression *Buh Buh Huh*, which orthographically represents the pronunciation of the acronym (9).

(5) **Bell Bearing Hunter** is the true final boss of this game
(chatter in LilAggy's stream 2478131530)

(6) If I was to choose a seed for my runs, I would avoid the wolf and **bell bearing** kkkkk my fights are always longer than I want
(chatter in nuclearpastatom's stream 2514398400)

(7) the **bell hunter** SUCKS
(chatter in LilAggy's stream 2477320404)

(8) You know, thinking about it, **BBH** would be a fun playable character in this with all his Darth Vader moves.
(chatter in itzCBD's stream 2516397731)

(9) **buh buh HUH** isn't a bad fight solo
(chatter in Captain_domo's stream 2482368557)

In the corpus, the most frequent variants are the official one (*Bell Bearing Hunter*), the acronym

[3]We also identified a whole range of variants that occur only once or twice (eg. *Barry the bell bearing meanie, Bell bearing Kunter, Chad bearing hunter* among others). These variants, although humorous, are unfortunately too rare to include in a quantitative analysis.

BBH and the other "acronym" *BUH BUH HUH*. The distribution of these variants is shown in Table 3.

Variant	Occurrences
Bell Bearing Hunter	289 (25.3%)
BBH	568 (49.8%)
BUH BUH HUH	108 (9.4%)
Other variants	176 (15.4%)

Table 3: Variation referring to the Bell Bearing Hunter

Table 3 shows that the acronym *BBH* is the most frequent (around 50% of occurrences), even more than the official name, which is used in only a quarter of references to the enemy in the corpus. The 'pronounced' acronym *BUH BUH HUH* is used in almost 10% of the cases.

As with the references to *Nightreign*, looking at the data at the level of individual streamers' chats is not very helpful. As Figures 3 and 4 shows, no variant is limited to the community of a single streamer.

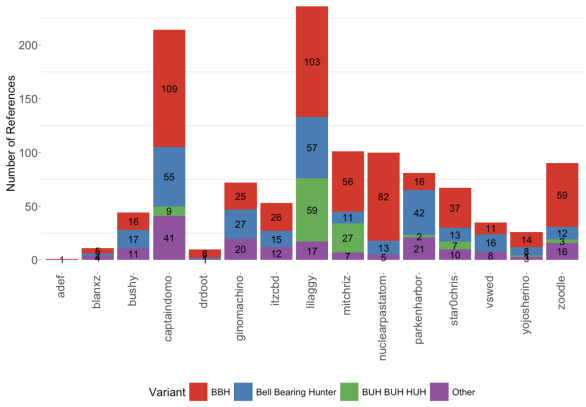

Figure 3: Bell Bearing Hunter vs BBH vs BUH BUH HUH by streamer (counts)

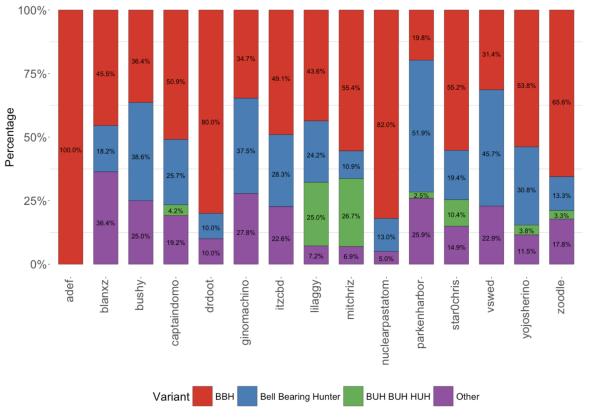

Figure 4: Bell Bearing Hunter vs BBH vs BUH BUH HUH by streamer (percentages)

However, in Figure 4, we can notice that the group of streamers whose chatters use *BUH BUH HUH* is quite similar to the chats where *NTR* is found. Both of these minority variants are found in the chats of LilAggy, Captain_domo, star0chris, parkenharbor and Mitchriz. In the next section, we will lay out a method for identifying more precisely what these streamers have in common.

4. Identifying meso-level communities on Twitch

We argue that the problem for studying linguistic variation with treating each streamer's community as a separate community of practice is that there is significant overlap in the audiences of the streamers. The streamers are all professional Elden Ring players who often collaborate with one another (for example, in the 2024 Bingo Brawlers tournament), so it is not surprising that a viewer who enjoys Elden Ring gameplay would view more than one streamer's channel. However, chatters' attentions are not spread out equally across all 15 streamers. For example, in the corpus, there are 3150 unique chatters that post in both LilAggy and Captain_domo's chats; however, there are only 709 chatters who contribute both to LilAggy and itzCBD's chat. This is despite Captain_domo and itzCBD having roughly the same number of followers, as shown in Table 1.

We can be more precise about how similar or different the chatting audience of the 15 streamers are through the use of the *Jaccard coefficient*, a similarity measure widely used in machine learning and various areas of quantitative and computational linguistics (see Tan et al. (2006)). For pairs of streamers, we calculated their Jaccard coefficient through taking the cardinality of the intersection of the sets of unique chatters for each streamer, and dividing this cardinality by the cardinality of the union of the two sets of unique chatters. In this way, Jaccard coefficients range from 0 (sets of chatters are completely disjoint) and 1 (sets of chatters are identical)[4].

Figure 5 shows a heatmap displaying the Jaccard coefficients measuring the chatter overlaps for each pair of streamers in the corpus. In line with the discussion above, we can observe that LilAggy and Captain_domo have a larger Jaccard coefficient than LilAggy and itzCBD (0.13 vs 0.03).

Figure 5 shows that the overlap between streamers' participating audiences is not enormous: the highest Jaccard coefficient is 0.24. Nevertheless, 0.24 does indicate that, for some pairs of streamers,

[4]The calculation of the Jaccard coefficients was done in R, and the code is available at https://osf.io/ct3zr/overview?view_only=395154716b8c4d5c9cbb955755a99b8c.

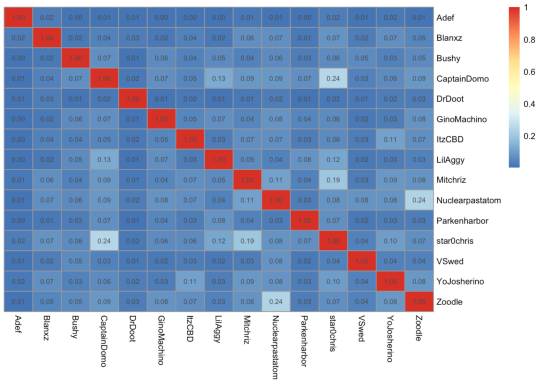

Figure 5: Heatmap of Jaccard coefficients for 15 streamers

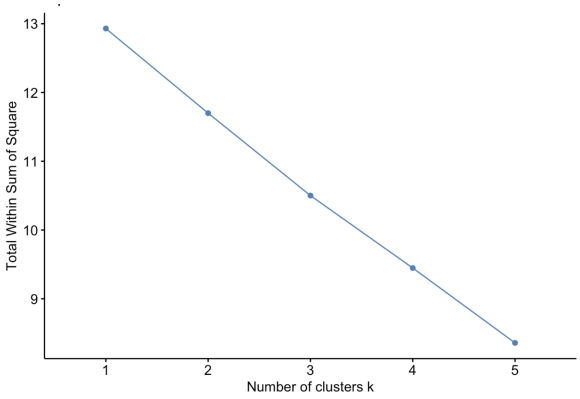

Figure 6: Total sum of squares for $k = 1$ to $k = 5$

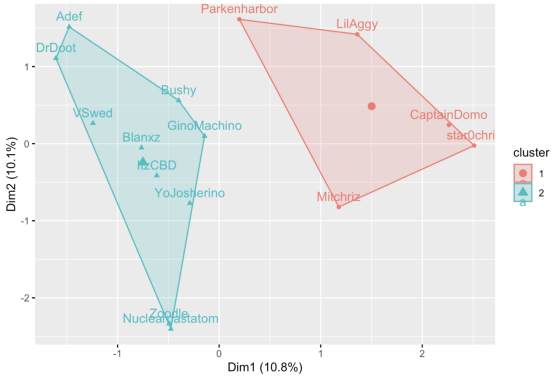

Figure 7: k-means clustering, for $k = 2$

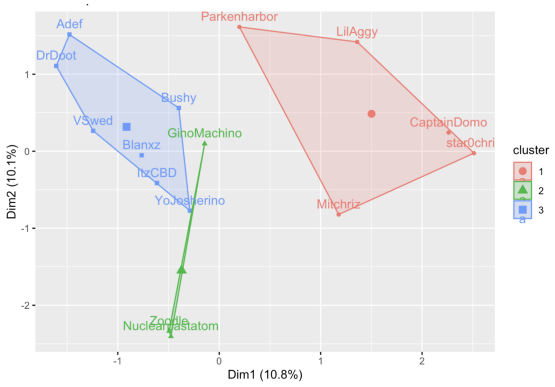

Figure 8: k-means clustering, for $k = 3$

a non-trivial part of their participating audience is shared. We would therefore like some way to group multiple streamers together based on this shared audience, to then see whether these new communities of practice are better predictors of linguistic variation.

For this purpose, we employed one of the most common clustering algorithms used in linguistics (see Levshina (2015)): *k-means clustering* (MacQueen, 1967). K-means clustering aims to partition (in this case) 15 streamers into k groups in a way that maximizes (in this case) Jaccard similarities between members of groups and minimizes similarities between members of different groups. The k-means clustering presented in this paper was done in R (with the *cluster* package (Maechler et al., 2026)), and the visualizations were also done in R (with the *factoextra* package (Kassambara and Mundt, 2020)).

Since we are using clustering as an exploratory technique, the difficult question of how many clusters are desired arises. As discussed above, all the Jaccard coefficients are quite similar, so we know that there are no isolated close-knit subcommunities within the group of 15 streamers being studied. This fact is reflected in the clustering results: Figure 6 plots the total within sum of squares value for results of the k-means algorithm (calculated using the *fviz_nbclust(·)* function from the *factoextra* package in R), for 1 to 5 clusters. In this figure, there is no “elbow", i.e. place where adding an extra cluster improves the fit of the model less than any other, at least for few numbers of clusters.

Nevertheless, we see that the algorithm consistently distinguishes between streamers like star0chris, Captain_domo and Mitchriz on the one hand, and nuclearpastatom and Zoodle on the other, as shown by the 2 cluster and 3 cluster visualizations in Figures 7 and 8.

In the next section, we will show that, even if we take a coarse-grained clustering, such as $k = 3$, we can arrive at a better model of ludolexical variation than if we simply consider either the whole community (15 channels) or each streamer’s channel independently.

5. Back to ludolexical variation

As shown in Figure 8, the three groups of streamers that were identified based on overlapping chatters are **Cluster A:** Parkenharbor, LilAggy, Captain_domo, star0chris and Mitchriz; **Cluster B:** Adef, Bushy, DrDoot, Vswed, Blanxz, itzCBD and YoJosherino; and **Cluster C:** Zoodle, nuclearpas-

tatom and Ginomachino.

Figure 9 shows the distribution of the variants *Nightreign, NR* and *NTR* to refer to the game by cluster. We can now observe that all the occurrences of *NTR* are found within the chats of streamers from cluster A.

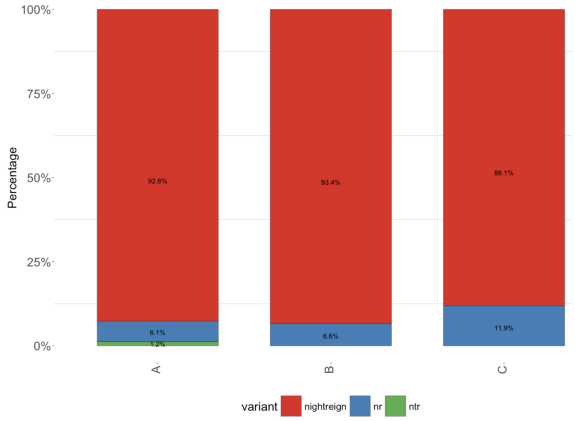

Figure 9: Nightreign vs NR vs NTR by cluster

We have some evidence that this pattern is driven at least in part by individuals who participate in multiple streams. For example, the first use of *NTR* in our corpus comes from a LilAggy stream, where a particular chatter, call them *Chatter A*, writes *NTR* as a variant for *NR* (10). The same chatter than participates in Captain_domo's stream a couple days later, using the new *NTR* variant (11). As example (11) shows, the innovative variant is remarked on by other Captain_domo chatters, who draw attention to the fact that *NTR* also happens to be an acronym for a genre of manga, *Netorare*, which is centered around infidelity and the dark feelings that ensue.

(10) idk about you guys, but ER **NTR** is at least and 8/10 for me
(Chatter A in LilAggy's stream 2482675698)

(11) **Chatter A:** I'm playing **NTR** on the other monitor and the double audio is really confusing
Chatter B : @[Chatter A] i wouldn't abbreviate nightreign to **ntr** KEKW
Chatter C: @[Chatter A] HANK, DONT ABBREVIATE NIGHTREIGN. HAAANK
(Captain_domo's stream 2492160644)

The NTR variant is then spread to Parkenhabor's chat, when Parkenharbor audience members observe LilAggy chatters using it, during a Nightreign collaboration a few days later (12)[5].

(12) why the fuck are they talking about **NTR** over there
(chatter in Parkenharbor's stream 2530353580)

In this way, we see the development of a linguistically relevant "meso-level" community of practice centered around Elden Ring gameplay, i.e. communities that are the result of (relatively loose) contact between smaller Twitch communities.

We find a similar pattern when it comes to referring to the Bell Bearing Hunter: as Figure 10 shows, *BUH BUH HUH* is significantly[6] more frequent in chats of streamers of the A cluster.

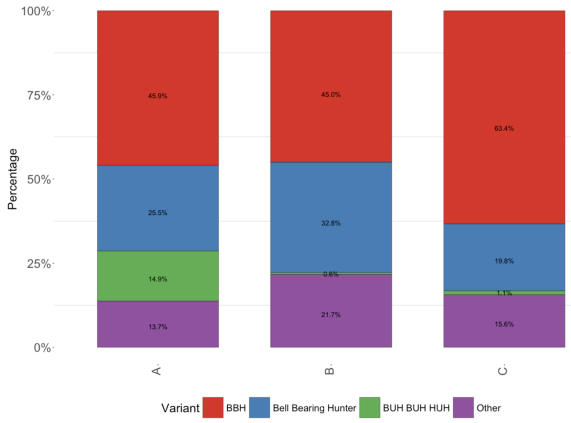

Figure 10: Bell Bearing Hunter vs BBH vs BUH BUH HUH by cluster

And it appears that a similar mechanism to the spread of *NTR* drives this pattern: individuals from one chat will bring variants they learn from that chat to another chat, which then starts using them. For example, a chatter in LilAggy's stream, call them *Chatter D*, uses the BUH BUH HUH variant when the Bell Bearing Hunter comes on screen. As example (13) shows, this is a common practice in LilAggy's stream. Chatter D then goes to Mitchriz's stream and informs Mitchriz's audience of the "correct pronunciation" of BBH.

(13) Chatter P: Buhh Buhh HUHH
Chatter Q: super tanky boi
Chatter D: Buhh Buhh HUHH
Chatter R: Buhh Buhh HUHH
(LilAggy stream 2501295308)

(14) it's pronounced Buh Buh HUH
(Chatter D, in Mitchriz stream 2518225917)

[5]We note here that clearly metalinguistic uses of *NTR* and *Nightreign* were excluded from the quantitative study. Thus, the data in Table 2 only counts actual references to the video game, not to the expression itself or to Netorare.

[6]In statistical analysis (generalized linear mixed effects models, using the *lm4e* R package (Bates et al., 2015), predicting the use of the *BUH BUH HUH* variant vs another variant, with cluster as fixed effet and streamID as random effect) we find that *BUH BUH HUH* is used significantly less in chats of cluster B ($p < 0.002$) and significantly less in chats of cluster C ($p < 0.001$), compared to the chats in cluster A.

With these two examples, we can see the development of a shared repertoire of linguistic practices (referring to Nightreign as *NTR* and the Bell Bearing Hunter as *BUH BUH HUH*) across communities associated with multiple Twitch channels : those that have many chatters in common.

6. Conclusion

This paper argued that, when it comes to modeling language variation and change on the video game streaming platform Twitch, it is necessary to consider “meso-level" communities of practice, i.e. ones that are smaller than the full video game community, yet larger than communities associated with individual channels. We provided a method, Jaccard coefficients + k-means clustering, which, we argued, could help identify these intermediary communities. We illustrated our proposals through a study of quantitative ludolexical variation in a corpus of chat transcripts from 15 streamers of “The Elden Ring Community". We showed that lexical innovations that are created in the community associated with a channel can be quickly propagated to the communities of other channels when overlap between the two communities is high, or when the two communities come into contact (for example through streamer collaborations). Our results thus provide new linguistic evidence for what Cochran et al. (2023) calls “nested communities of practice", CofPs that exist within a larger gaming community of practice, yet may also contain other CofPs.

7. Limitations and ethics statement

The results in this paper raise a number of questions that we are currently unfortunately unable to answer. For example, we focused on the linguistic productions of chatters, neglecting those of the streamers. The reason for this focus was practicality: the chat transcripts were easily obtained through an application; whereas, obtaining a transcript of the streamers' linguistic productions would be extremely difficult and time consuming, since the stream videos are often 5-10 hours long and automatic transcription of technical gameplay language is mediocre. However, the streamers' language is clearly relevant to the development of shared lingusitic repertoires, so lack of information about the streamers' behavior is a clear limitation of this study.

Another question raised by our results concerns the precise method for discovering intermediary CofPs. As Figure 6 shows, k-means clustering for $k < 6$ does not provide a good fit for our data. Even though we argued that this method was sufficient to shed light on linguistic variation in our corpus, we might wonder whether other methods of identifying subgroups of streamers or chatters would yield better results. We therefore leave the exploration of this possiblity to future research.

Ethics statement: The study reported in this paper used only publicly available data from individuals who, at the time of their participation in public Twitch streams, gave their consent to have their productions recorded and broadcast. No artificial intelligence tools were used at any point in the data, analysis or writing of this paper.

8. Acknowledgments

The author would like to thank Olivier Bonami and two SoCon reviewers for helpful comments and suggestions. This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 850539).

9. Bibliographical References

Benjamin Anderson. 1983. *Imagined Communities: Reflections on the Origins and Spread of Nationalism*. Verso.

Jo Angouri. 2015. Online communities and communities of practice. In *The Routledge handbook of language and digital communication*, pages 323–338. Routledge.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Mary Bucholtz. 1999. “why be normal?”: Language and identity practices in a community of nerd girls. *Language in society*, 28(2):203–223.

Andrew F Cochran, Stephen T Slota, and Michael F Young. 2023. League of legends. *The Literacies of the Esports Ecosystem*, 5:93.

Penelope Eckert. 2000. *Linguistic Variation as Social Practice*. Blackwell, Oxford.

Penelope Eckert. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41(1):87–100.

Penelope Eckert and Sally McConnell-Ginet. 1992. Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology*, pages 461–490.

Astrid Ensslin. 2012. *The language of gaming*. Bloomsbury Publishing.

Sage L Graham. 2019. A wink and a nod: The role of emojis in forming digital communities. *Multilingua*, 38(4):377–400.

Daniel Gros, Anna Hackenholt, Piotr Zawadzki, and Brigitta Wanner. 2018. Interactions of twitch users and their usage behavior. In *International Conference on Social Computing and Social Media*, pages 201–213. Springer.

William A Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on Twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1315–1324.

Janet Holmes, Maria Stubbe, and Bernadette Vine. 1999. Constructing professional identity:"doing power" in policy units. *Talk, work and institutional order: Discourse in medical, mediation and management settings*, 1:351–385.

Joshua Boyd Iorio. 2010. *Explaining orthographic variation in a virtual community: Linguistic, social, and contextual factors*. Ph.D. thesis, University of Texas at Austin.

Salwa Raudhatul Jannah and Tatan Tawami. 2024. Morphological shortening in mobile legend video game. *Mahadaya: Jurnal Bahasa, Sastra, dan Budaya*, 4(2):271–288.

Alboukadel Kassambara and Fabian Mundt. 2020. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.999.

William Labov. 1966. *The social stratification of English in New York city*. Cambridge University Press.

William Labov. 1973. *Sociolinguistic patterns*. 4. University of Pennsylvania press.

Sven Leuckert and Martin Leuckert. 2020. Towards a digital sociolinguistics: Communities of practice on reddit. In *Corpus approaches to social media*, pages 15–40. John Benjamins Publishing Company.

Natalia Levshina. 2015. *How to do linguistics with R*. John Benjamins Publishing Company.

J MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings ofthe 5th Berkeley Symposium on Mathematical Statisticsand Probability*, volume 1, pages 281–297.

Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2026. *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.8.2 — For new features, see the 'NEWS' and the 'Changelog' file in the package source).

Rick Marlatt. 2020. Capitalizing on the craze of fortnite: Toward a conceptual framework for understanding how gamers construct communities of practice. *Journal of Education*, 200(1):3–11.

Miriam Meyerhoff. 2013. Communities of practice. *The handbook of language variation and change*, pages 428–447.

Gillian Sankoff and Henrietta Cedergren. 1972. Sociolinguistic research on french in montreal. *Language in Society*, 1(1):173–174.

Pang-Ning Tan, Michael Steinbach, Vipin Kumar, et al. 2006. Cluster analysis: basic concepts and algorithms. *Introduction to data mining*, 8:487–568.

JTA van der Aa. 2021. "it'sa first!": A case study on identity expression through language on twitch in a game of teamfight tactics. B.S. thesis, Utrecht University.

Etienne Wenger. 1998. *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Walt Wolfram. 1969. *Black/White Speech Differences Revisited: A Preliminary Report*. Center for Applied Linguistics, Washington DC.

Frederic Zähres. 2019. " pentakill inc."-a case study of shortening processes in ludolects. In *Communication forms and communicative practices. New perspectives on communication forms, affordances and what users make of them*, volume 15. Peter Lang.