Vocabulary Circulation in French-speaking Video Game Streaming on *Twitch.tv*: *How to Model and Make Interpretable the Intersection between Network Analysis and NLP?*

Fanny Barnabé
Epitech, MNSHS Laboratory
E-mail: fanny.barnabe@epitech.eu

Nicolas Bourgeois
Epitech, MNSHS Laboratory
E-mail: nicolas.bourgeois@epitech.eu

Abstract

Video game live-streaming is a form of performance that continues to grow in popularity and that exerts considerable influence not only on the ways in which video game culture is consumed, but also on the forms of participation in that culture, particularly through the various forms of "secondary play" (Delbouille 2018). Because of this impact and popularity, live-streaming already benefits from an important bibliography in the field of game studies and has its own methodologies (Recktenwald 2017). In particular, this object has been studied in relation to competitive gaming practices and e-sports (see Taylor 2012 and 2018), through the concept of game spectatorship (Cheung & Huang 2011), or in the fields of sociology, ethnography or marketing (see Gandolfi 2016; Hamilton, Garretson & Kerne 2014; Kaytoue *et al.* 2012; Deng *et al.* 2015; Edge 2013; Johnson & Woodcock 2019; Pellicone & Ahn 2017) among others (for an expanded state of the art on the topic, see Harpstead *et al.* 2019).

However, while it uses video game as a medium, this cultural practice also relies heavily on the production of spoken and written discourses. On the one hand, the streamers' utterances federate their communities through the construction of an idiolect (using slogans, running gags, catchphrases), more or less fed and crossed by other lexical fields (coming from the games played or the streaming platform, in particular). On the other hand, the chat's written interventions repeat, appropriate and make these language elements evolve. A few works have begun to focus on these discourses, especially for their potentially toxic dimension (see Poyane 2018 and 2019; Mihailova 2020), or to underline the massive and

cacophonous character of viewers' interventions in the chat, sometimes compared to the "roar of a crowd in a stadium" or to a "waterfall of text" (Hamilton *et al.* 2014: 1320).

As an extension of these first approaches, this presentation proposes to study the circulation of vocabulary between several communities of French-speaking streaming viewers on *Twitch.tv*, by combining (following the model of the work of Bouveyron, Latouche & Zreik, 2016) network analysis and natural language processing of messages posted in the chat. Specifically, we will build on the work of Nicholas Bouchaib[1], who used network analysis to map the various French-speaking streaming communities on Twitch during January 2022. By measuring viewers traffic on different channels, he produced several graphs (where each point represents an account and each link the action of watching a streamer) showing which channels attract the largest audience and which channels are connected by a large number of viewers.

Figure 1 - Cartography of the "multi-gaming" streaming communities realized by Nicholas Bouchaib

---

[1] https://twitter.com/nicolasbchb/status/1491113153396375552?s=20&t=KYeNJz6YzuvRcvnDJAR00w

Based on these first visualizations, we will select a corpus of chat logs from the 9 main channels emerging from the "multigaming" subcategory of the map (see Figure 1): Domingo, Zerator, Ponce, Antoine Daniel, MisterMv, Joueur du Grenier, Étoiles, Le Stream and Xari. This category gathers "variety streamers" (Taylor 2018), i.e. streamers who use various video games in order to offer content mainly aimed at humor and entertainment (rather than competition). Thanks to the tool *TwitchDownloader*[2], we will download, for each considered channel, one week of chat logs in the form of text files including the messages posted, their time stamp and the pseudonyms of their authors (this should represent several hundreds of thousands of entries in the chat for each channel, equivalent to several million words).

By exploiting this corpus with the help of natural language processing tools, this paper will aim at answering two questions in parallel. On the one hand, we will try to determine to what extent communities defined by the network analysis can also be identified by the sharing of the same vocabulary: how do discursive communities form on *Twitch.tv*? Around what type of vocabulary do they form or break down? What are the elements that determine the appearance of one lexical field rather than another? What is the share of influence of the streamer or of the game played, and that of the platform *Twitch.tv*?

On the other hand, we will try to answer a more methodological question related to digital humanities (which can to some extent be transposed to the topic of character networks): how to represent such a large number of superimposed data in an interpretable and efficient model? One of the shortcomings of the visualizations proposed by Nicolas Bouchaib is indeed that the density of connections within the network makes it difficult to read the graphs and invites us to question what really provides information in this context, and how to preserve the ratio between the quantity of information transmitted and the complexity of the infographics (Tufte 2020). In our case, the representation will require even more a semiotic reflection, since it will be necessary to make visible the exchanges and evolution of the vocabulary used by the audience in addition to the particularly dense network of actors. Through this case study, we thus wish to shed light on the modes of operation of linguistic exchanges specific to video game streaming on *Twitch.tv*, but also to propose methods facilitating the study of the circulation of vocabulary between multiple entities.

References

Charles Bouveyron, Pierre Latouche, Rawya Zreik. 2016. The stochastic topic block model for the clustering of vertices in networks with textual edges. *Statistics and Computing*, 28: 11–31. https://doi.org/10.1007/s11222-016-9713-7.

Gifford Cheung & Jeff Huang. 2011. Starcraft from the stands: Understanding the game spectator. In *CHI'11 Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*. ACM Publications, pages 763-772.

Julie Delbouille. 2018. Jouer à travers l'autre : jeu « secondaire », jeu par procuration et médiation du plaisir ludique. Paper presented at the conference *Appropriation, pratiques et usages: en/jeux. Montréal*. http://hdl.handle.net/2268/223638.

Jie Deng, Félix Cuadrado, Gareth Tyson & Steve Uhlig. 2015. Behind the Game: Exploring the Twitch Streaming Platform. Paper presented at the conference *NetGames 2015. The 14th International Workshop on Network and Systems Support for Games*. Zagreb. https://eecs.qmul.ac.uk/~fcuadrado/papers/netgames15-twitch.pdf.

Nathan Edge. 2013. Evolution of the Gaming Experience: Live Video Streaming and the Emergence of a New Web Community. *The Elon Journal of Undergraduate Research in Communications*, 4(2).

Enrico Gandolfi. 2016. To watch or to play, it is in the game: The game culture on Twitch.tv among performers, plays and audiences. *Journal of Gaming & Virtual Worlds*, 8(1): 63-82.

William A. Hamilton, Oliver Garretson & A. 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*. ACM Publications, pages 1315-1324. http://ecologylab.cse.tamu.edu/research/publications/streamingOnTwitch.pdf.

---

[2] https://github.com/lay295/TwitchDownloader

Erik Harpstead, Juan Sebastian Rios, Joseph Seering & Jessica Hammer. 2019. Toward a Twitch Research Toolkit: A Systematic Review of Approaches to Research on Game Streaming. In *CHI PLAY '19: Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM Publications, pages 111–119. https://doi.org/10.1145/3311350.3347149.

Mark R. Johnson & Jamie Woodcock. 2019. The Impacts of Live Streaming and Twitch.tv on the Video Game Industry. *Media, Culture and Society*, 41(5): 670-68. https://doi.org/10.1177/0163443718818363.

Mehdi Kaytoue, Arlei Silva, Loïc Cerf, Wagner Meira Jr. & Chedy Raïssi. 2012. Watch me playing, I am a professional: a first study on video game live streaming. In *Proceedings of the 21st International Conference Companion on World Wide Web*. ACM Publications, pages 1181-1188. http://homepages.dcc.ufmg.br/~kaytoue/KSCMR-MSND12.pdf.

Teodora Mihailova. 2020. Navigating ambiguous negativity: A case study of Twitch.tv live chats. *New Media & Society*, https://doi.org/10.1177/1461444820978999.

Anthony J. Pellicone & June Ahn. 2017. The Game of Performing Play: Understanding Streaming as Cultural Production. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM Publications, pages 4863–4874. http://library.usc.edu.ph/ACM/CHI%202017/1proc/p4863.pdf.

Roman Poyane. 2018. Toxic communication during streams on Twitch.tv. The case of Dota 2. In *Proceedings of the 22nd International Academic Mindtrek Conference (Mindtrek '18)*. Association for Computing Machinery, pages 262–265. https://doi.org/10.1145/3275116.3275152.

Roman Poyane. 2019. Toxic Communication on Twitch.tv. Effect of a Streamer. In Daniel A. Alexandrov et al., eds., *Digital Transformation and Global Society. DTGS 2019*. Springer, pages 414-421.

Daniel Recktenwald. 2017. Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, 115: 68-81. http://dx.doi.org/10.1016/j.pragma.2017.01.013.

T. L. Taylor. 2012. *Raising the stakes: E-sports and the professionalization of computer gaming*. MIT Press, Cambridge.

T. L. Taylor. 2018. *Watch Me Play: Twitch and the Rise of Game Live Streaming*. Princeton University Press, Princeton.

Edward R. Tufte. 2020. Seeing With Fresh Eyes: Meaning, Space, Data, Truth. Graphics Press, Cheshire.