

**En écrivant comme Beckett:
A dialogue between generative AI and genetic criticism**

Jonathan Hoebeke, KU Leuven (student, Advanced Interuniversity MA in Literary Studies)

Category: short paper

In 2019, Dirk Van Hulle imagined the possibility of a writing bot which would make use of the insights and material available thanks to genetic criticism—a discipline which uses draft material to examine writing processes (Van Hulle 2019). The award-winning Beckett Digital Manuscript Project (BDMP) offers digital facsimiles and transcriptions of Samuel Beckett’s manuscripts, genetic editions of his texts based on them, and a digital reconstruction of his personal library. Beckett’s oeuvre lends itself particularly well to a world trying to come to grips with generative AI, as it anticipates some of this technology’s central concerns. Described as being written in a “proto-computer-language” (Kenner 1987, 96), it explores themes relevant to the emergence of computer-generated text, such as the limits of language, the (im)possibility of silence, or the quest for a self. This study therefore aims to put Van Hulle’s thought experiment into practice by creating a ‘BeckettBot’ trained on material derived from the BDMP, thus bringing together genetic criticism and generative AI in a dialogue between two disciplines which have much to say about authorship and creativity.

The BDMP provides two kinds of resource (endogenetic and exogenetic, i.e. the successive drafts forming the ‘inside’ of the writing process of a text versus those ‘outside’ sources originally external to the text but relevant to its writing) that are useful for this ongoing project. First, it gives access to genetic editions of Beckett’s writings. These editions reconstruct the author’s writing process, allowing for a deeper understanding of how his texts took shape. Second, the Beckett Digital Library (BDL), described in an accompanying monograph (Van Hulle and Nixon 2013), offers a digital version of the author’s personal library and makes note of any marginalia or other reading traces. Instead of simply relying on an existing model’s pre-trained masses of textual data from all kinds of (unknown) sources, this virtual library thus enables the construction of a training corpus that more carefully emulates Beckett’s actual reading practice. Together, these resources are an opportunity to go beyond machine-learning approaches which only consider final, published texts in their training material. By including the insights into the creative process offered by genetic research and the intertextual traces revealed by the virtual library, this study hopes to contribute a new perspective to the current wave of so-called ‘generative AI’, while also being aware that this very term can be misleading when applied to what some have termed “stochastic parrots” (Bender et al. 2021), which are arguably “neither artificial nor intelligent” (Crawford 2021, 8).

One of the striking aspects of Beckett’s writing is its bilingual nature, as Beckett systematically wrote (and self-translated) his texts in English and in French. Unsurprisingly, his library also contains many books in both languages (as well as others, mainly in German and Italian). This project therefore not only looks to expand standard generative practices vertically, by digging deeper than the surface-level of authors’ published texts, but also horizontally, by experimenting with the multilingual potential embedded in Beckett’s writing practice.

The goal of this research, then, is to create a ‘BeckettBot’ (indeed, multiple bots—a vital point is that there can be no such thing as *the* BeckettBot) by fine-tuning a large language model (LLM) on the material described above, which is then able to read and produce text in a ‘Beckett-like’ manner. LLMs like OpenAI’s various iterations of their Generative Pre-Trained Transformer (GPT) have shown great promise in generating texts which come (uncannily) close to what a human might produce. Both LLMs’ ability to convince sometimes and various failures at other times make them excellent research objects for humanities scholars interested in the creative process and (the limits of) language. In concrete terms, for the purposes of this presentation, and as an initial step towards the broader aims of this project, the proposed model will be trained on the bilingual endogenous material available for this project. In a first move, I will showcase a BeckettBot trained solely on Samuel Beckett’s published texts. This is already an update to previous computer-generated extensions of Beckett’s writing, such as the *Megawatt* project (Montfort 2014), which used a human-written, deterministic, rule-based program to generate a new version of Beckett’s novel *Watt*, rather than today’s self-supervised LLMs. At the same time, this way of superficially recreating a writer’s voice is by now quickly becoming a familiar and fairly straightforward move. More interesting is the next step: the introduction of another BeckettBot, trained on the genetic editions provided by the BDMP. Beyond the stylistic question of whether a generated text ‘feels Beckettian’, this approach allows us to observe to what extent we can now emulate Beckett’s writing process, for example by comparing known aspects of his revision habits (such as more deleting than adding) with the new text the bot generates when given a draft version. Both BeckettBots will come in two variants: one trained exclusively on his English texts and one exclusively on the French ones, enabling a comparative, multilingual analysis.

Coupling existing LLMs with the genetic dossier of an author like Beckett, whose writing process and output already heavily undermine any notion of a straightforward text and imaginative process, holds great promise for a mutually enriching encounter. Active experimentation—including and perhaps especially by humanities (and specifically literary) scholars—is needed to gain a better understanding of these technologies and their limitations. This study wishes to be part of a dialogue between disciplines which, in typical Beckettian fashion, in some ways has just gotten started, in others is as old as time, and which must—perhaps forever—go on.

References

- Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FaccT '21)*, 2021, pp. 610-23, doi:10.1145/3442188.3445922.
- Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.
- Kenner, Hugh. "Beckett Thinking." *The Mechanic Muse*. Oxford University Press, 1987, pp. 83-105.
- Montfort, Nick. *Megawatt*. 2014. *nickm.com*, www.nickm.com/poems/megawatt.pdf. PDF file. Accessed 1 Feb. 2023.
- Van Hulle, Dirk. "Artificial imagination, imagine: new developments in digital scholarly editing." *International Journal of Digital Humanities*, vol. 1, no. 2, July 2019, pp. 137-40, doi:10.1007/S42803-019-00020-W.
- Van Hulle, Dirk, and Mark Nixon. *Samuel Beckett's Library*. Cambridge University Press, 2013.