



**HAL**  
open science

## Exploring Combinatorial Methods to Produce Sonnets: An Overview of the Oupoco Project

Frédérique Mélanie-Becquet, Clément Plancq, Claude Grunspan, Mylène Maignant, Matthieu Raffard, Mathilde Roussel, Fiammetta Ghedini, Thierry Poibeau

### ► To cite this version:

Frédérique Mélanie-Becquet, Clément Plancq, Claude Grunspan, Mylène Maignant, Matthieu Raffard, et al.. Exploring Combinatorial Methods to Produce Sonnets: An Overview of the Oupoco Project. Digital Humanities Quarterly, 2024, 18 (1). hal-04855286

**HAL Id: hal-04855286**

**<https://hal.science/hal-04855286v1>**

Submitted on 24 Dec 2024


**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Distributed under a Creative Commons Attribution 4.0 International License

## Exploring Combinatorial Methods to Produce Sonnets: An Overview of the Oupoco Project

Frédérique Mélanie-Becquet <frederique\_dot\_melanie\_at\_ens\_dot\_psl\_dot\_eu>, LATTICE (CNRS & ENS/PSL & Univ. Sorbonne Nouvelle)  <https://orcid.org/0000-0002-8216-9437>

Clément Plancq <clement\_dot\_plancq\_at\_univ-tours\_dot\_fr>, LATTICE (CNRS & ENS/PSL & Univ. Sorbonne Nouvelle)  <https://orcid.org/0000-0001-6189-0703>

Claude Grunspan <grunspanclauda\_at\_gmail\_dot\_com>, LATTICE (CNRS & ENS/PSL & Univ. Sorbonne Nouvelle)


Mylène Mignant <mignant\_dot\_mylene\_at\_gmail\_dot\_com>, LATTICE (CNRS & ENS/PSL & Univ. Sorbonne Nouvelle)  <https://orcid.org/0000-0001-8046-7783>

Matthieu Raffard <raffard\_dot\_rousseau\_at\_gmail\_dot\_com>, Atelier Raffard-Roussel

Mathilde Roussel <raffard\_dot\_rousseau\_at\_gmail\_dot\_com>, Atelier Raffard-Roussel

Fiammetta Ghedini, RIVA Illustrations

Thierry Poibeau <thierry\_dot\_poibeau\_at\_ens\_dot\_psl\_dot\_eu>, LATTICE (CNRS & ENS/PSL & Univ. Sorbonne Nouvelle)

 <https://orcid.org/0000-0003-3669-4051>

### Abstract

In this paper, we describe Oupoco (l'Ouvroir de Poésie Combinatoire), a system producing new sonnets by recombining lines of poetry from existing sonnets, following an idea that Queneau described in his book *Cent Mille Millions de poèmes (A Hundred Thousand Billion Poems, 1961)*. We first give the rationale of the project and review past experiments in poetry generation using combinatorial methods. We then demonstrate different outputs of our implementation (a Web site, a Twitter bot and a specifically developed device, called the *Boîte à poésie*) based on a corpus of 19th century French poetry. We describe how this project was an opportunity to work with artists and reach a new audience through the *Boîte à poésie*, and also through a video clip that frequently served as an introduction to the project. Our goal is to revive people's interest in poetry by giving access to automatically produced sonnets through original and entertaining channels and devices.

## 1 Introduction: Combinatorial Literature, from Queneau to Oupoco

The Oupoco project is a project exploring combinatorial methods to produce poetry by recombining existing lines of poetry. The starting point of the project was the book by Raymond Queneau, *Cent Mille Millions de poèmes (A Hundred Thousand Billion Poems, in English)* [Queneau 1961]. 1

This book is composed of ten sonnets, each line being printed on a separate strip of paper (Figure 1), allowing the reader, by freely combining the different lines, to potentially produce and read a hundred thousand billion different sonnets. This combinatorial approach (the main characteristic of the book) encourages the reader to play with the meanings of the lines, with the various language registers, or simply to have fun with the book. It is this stimulating relationship between poetry and constraints that we wanted to reproduce and that was the mainspring of our project. 2



Figure 1. *Cent Mille Millions de poèmes* by Queneau, both a book and an artistic creation

Queneau's book seems an ideal candidate to be transposed to a computer, since a machine can very easily combine lines and produce a comprehensive list of all the possible poems. However, Queneau's work is still under copyright, which should normally prevent one from putting the sonnets contained in his book online, although numerous implementations can in fact be found on the Internet. Nevertheless, this is one of the reasons why, instead of working on Queneau's poems, we decided to focus on freely available ones, which also helped develop more original and challenging research. Hence, an important part of the project consisted in constituting a collection of 19th century French sonnets.

3

Beyond the availability of relevant collections of poetry, which facilitated the creation of our corpus, 19th century poetry also appeared as a fertile ground to reflect upon the sonnet and to work and play with various constraints. This means that the original poems first have to be analyzed so as to determine the rhyme, as well as other features (line length, topic of the sonnet, etc.). Our project thus deals with analysis, as much as with the generation of poetry. We will see that the project gave birth to various outputs, including artistic ones. Finally, we made several public demos, showing that such a project can have an impact on the general public as well as on children in school, in a learning perspective.

4

The structure of the paper is as follows. We will first examine some past work and give our motivations for this project (Section 2). We then describe the corpus (Section 3), the approach for rhyme analysis (Section 4), and the constraints on the generation process (Section 5). We conclude the technical part of the paper with an evaluation of the sonnets thus generated (Section 6). We then describe our collaboration with artists, to produce the *Boîte à Poésie*, and the Oupoco video clip (Section 7). This led to different demonstrations and some outreach initiatives with the public (Section 8). We conclude with a discussion on computers and creativity (Section 9).

5

## 2 Computational Combinatorial Poetry, from Queneau to Oupoco

In this section, we examine some past experiments to generate poetry with computers in the context of the Oulipo <sup>[1]</sup>. We then describe the rationale behind our project, inspired by Queneau but also largely departing from Oulipo's original ideas.

6

## 2.1 Transposing Queneau's Book to a Computer

As mentioned in the introduction, there have been various attempts to transpose Queneau's book to a computer. In fact, in 1961, immediately after the publication of the book, a first attempt was made by Dimitri Starynkevitch, a French computer scientist, working on a computer of the time (a CAB 500, to be precise, according to Bens (2005); see Figure 2 and [Starynkevitch 1990]. Baillehache (2021) reports Oulipo's answer to Starynkevitch's program: "We wished that M. Starynkevitch would explain the method used; we hoped that the choice of the verses was not left to randomness" [Bens 2005, 79]. As Baillehache explains, this answer "reflects mistrust against randomness"; moreover, it shows Oulipians' lack of interest, in 1961, in the process that would make it possible to generate poetry indefinitely and automatically, which may seem surprising at first glance.

7



**Figure 2.** A picture of the CAB 500 computer (designed by the French company SEA, around 1957, by Alice Recoque. [Starynkevitch 1990])

It is indeed not entirely clear why Queneau and Oulipo reacted in this way. Baillehache (2021), after Bloomfield and Campaignolle (2016), attributed it to Oulipo's aversion to randomness, seen as opposite to the idea of constraints that the group explored intensively in its literature.

8

While this is possible, we also think that the main goal of Queneau was not just to produce an infinite number of poems generated randomly. The book, with its strips of paper, is in fact more than a book, it is also a piece of art and an object conceived as a device to be manipulated by people. The whole idea is to assemble lines of verse, look at the result, try other possibilities and produce silly, funny, or just unexpected poems. The transposition to a computer in 1961, and most of the implementations one can find online nowadays, do not allow this: they produce poems, one at a time, randomly, without any possible interaction with the end-user or the reader.

9

Since the advent of the Web in the 1990s, many other implementations (transposing Queneau's book and ideas online) have been proposed over the years. But there were many attempts even before that. Leonardo/Olats and Philippe Bootz (2006) give an interesting survey of what they call *combinatorial generative literature* (*littérature générative combinatoire*, in French). The authors report that Paul Braffort, from Oulipo, developed a computer version of the book in 1975, for an exhibition (Europalia) in Brussels. Tibor Papp, another enthusiast of computer literature, who worked with Bootz in the 1980s, also developed in 1988 an electronic version of the book (Figure 3). Last but not least, a version was developed by Antoine Denize and Bernard Magné for the CD-ROM *Machines à écrire*, in which the text is enhanced with graphic effects that were not part of Queneau's original project (Figure 4). This CD-Rom, published by one of the most prestigious publishers in France (who is also the publisher of Queneau's work), had a wider audience

10

that the previous programs, which reached only a limited readership.

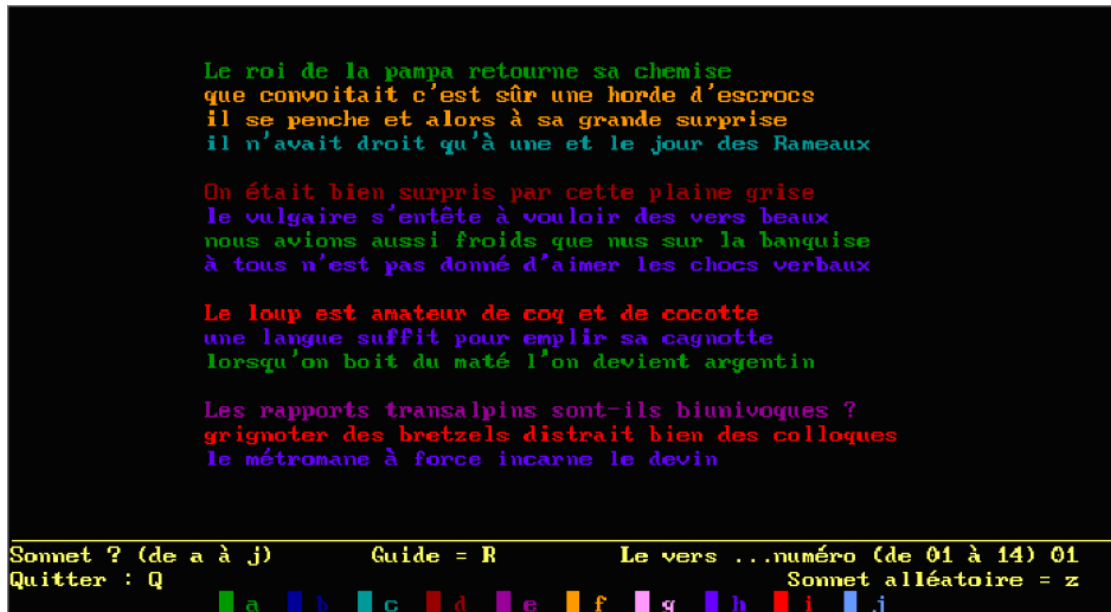


Figure 3. A screenshot of Tibor Papp's implementation, in 1988 (reproduced from [Leonardo/Olats and Bootz 2006] thanks to the kind authorization of Philippe Bootz).



Figure 4. Antoine Denize's version, enhanced with graphic effects, for the CD-Rom *Machines à écrire* for the French publisher Gallimard [Denize and Magné 1999], (reproduced from [Leonardo/Olats and Bootz 2006] thanks to the kind authorization of Philippe Bootz).

In the 1990s, a natural move consisted in implementing versions on the Web. A memorable event occurred in 1997, when a French computer science student was summoned to court by a descendant of Queneau, because he had put online a version replicating Queneau's book (thus breaching the copyright). The event attracted some attention, at least

in France, as one of the first cases sending someone to court for a violation of literary property on the Internet<sup>[2]</sup>. In the end, the court sentence was mild (as the program had already been removed from the Internet). Since then, the magnitude of the problem of online piracy (the practice of downloading and distributing copyrighted works digitally without permission) has downgraded this episode to a mere incident. To the best of our knowledge, nobody else since then has encountered problems with online versions of Queneau's book, even if they infringe copyright laws. Their readership is in any case extremely limited.

While Oulipo members initially showed some interest in computers, this did not last for very long, as we have seen above. To revive this topic, a specific group, called Alamo (*Atelier de Littérature Assistée par la Mathématique et les Ordinateurs*), comprising members of the Oulipo but also (and mostly) academics and computer scientists interested in the connections between literature and computers, was created in 1982. While this group developed interesting ideas and pieces of software, it remained independent of the Oulipo, and in turn the Oulipo did not show a great interest in the activities of this group. Moreover, the rapid obsolescence of computer programs made their production hard to maintain and showcase. More recently, Natalie Berkman (2017, 2022) tried to reproduce some of the early experiments of the Oulipo with computers and, more specifically, she tried to understand why the Oulipo quickly abandoned the idea of exploring this topic. Developing a program on a computer in the 1960s was complex (it involved preparing large sets of punch cards for example; it is no coincidence that the Alamo was created in 1982, with the advent of the personal computer). Moreover, it was not really possible at that time to represent the whole complexity of a language on a computer, and computers did not offer any real advantages compared to mathematical or formal models described on paper. As a result, in Berkman's words (2022), the Oulipo's members showed "a clear preference for abstract mathematical thought — patterns and structure — rather than the procedural tendencies of applied mathematics". But let's now come back to our own experiment.

12

## 2.2 Beyond *Cent Mille Milliards de poèmes*, our own experiment

In an interview with Georges Charbonnier (1962), Queneau explained that *potential literature* (*la littérature potentielle*) is "the search for forms, [...] for new structures which can then be used by writers in whatever way they like" [Charbonnier 1962, 140]<sup>[3]</sup>. Our own experiment here aims at reviving not only an interest in poetry but also aims at exploring new ways of writing, at a time when text is everywhere around us. For this, we also took inspiration from the notion of *uncreative writing*, as defined by Kenneth Goldsmith [Goldsmith 2011] [Goldsmith 2019].

13

*Uncreative Writing* challenges conventional notions of creativity and authorship in the realm of literature. Goldsmith's approach advocates for a departure from the traditional, romanticized image of the writer as a solitary genius crafting original, imaginative works. Instead, he promotes the idea that creativity can also be found in the act of appropriation and remixing, where writers repurpose existing texts, often without significant alteration, to create something new and thought-provoking. This notion blurs the lines between authorship, plagiarism, and creativity, emphasizing the power of the remix culture and the vast reservoir of existing texts as a source of inspiration. *Uncreative Writing* invites us to reconsider the boundaries of creativity in the digital age, where information is abundant and easily accessible, fostering a literary landscape that celebrates the art of recontextualization and the collaborative nature of language itself.

14

Another important point, especially if we place ourselves in the continuation of the Oulipo tradition, is the role of randomness (or chance) in generating texts. In Jacques Bens' account of the early Oulipo meetings, there is extensive discussion on this topic, and Queneau believed that writers should always be conscious of the constraints they are using and leave nothing to chance. In this sense, Oulipo wanted to affirm its departure from other literary movements, first and foremost surrealism. Our project aligns with this idea, in a way: we focus on sonnets, and the goal is to respect as much as possible the constraints of this specific poetic form. However, the notion of chance should not be confused with the notion of *potential literature* (*la littérature potentielle*): there is still an infinity of possible (potential) poems, even when taking into account all the constraints that apply [Fournel 1972] [Oulipo 1973] [Oulipo 1981] [Oulipo 2002].

15

Producing computational combinatorial poetry inspired by Raymond Queneau's *One Hundred Thousand Billion Poems* using publicly available poems opens up a large number of possibilities. By dissecting and reassembling publicly accessible verses through code, we can explore an infinity of poetic possibilities, taking into account constraints. Each

16

line and word are placed in a new context, forging connections and juxtapositions that transcend the individual poems' original contexts. Through this experiment, we pay homage to Queneau's legacy while showcasing the potential of technology to popularize poetry, enabling anyone to explore poetry and computational creativity.

### 3 The Corpus

The first part of this research consisted in gathering a corpus of French sonnets from the 19th and early 20th century. Although there are famous sonnets from the 16th and 17th century in French, we choose to focus mainly on the 19th century in order to build a more homogeneous corpus, both from a thematic and linguistic point of view (19th century French is more normalized than the language a few centuries before). We first used resources that were freely available on the Web, especially from the Gutenberg project and Wikisource.

The *Bibliothèque nationale de France* (BnF) has an even larger collection of poems since it is in charge of legal deposit for France. A large proportion of these books has been digitized and is available on the Gallica website (maintained by the BnF). The BnF was thus able to select and extract a large number of poetry books, which included various collections, anthologies and essays (all the data are in the XML Alto format). We selected the books with an OCR quality score above 98%, and used the BnF API to collect metadata about the books.

As these are complete books, not everything in them was relevant. Some books were simply irrelevant (e.g., critical essays that do not contain poetry), some parts had to be removed (preface, introduction, critical notes) and, even once we had reduced the text to poetry, we still needed to extract sonnets. In order to do this, several heuristics were used, thanks to Python scripts aiming at retrieving poems in the form: two quatrains followed by two tercets (the form of a sonnet). This heuristic (based simply on the analysis of space and empty lines between verses) proved quite robust and efficient, even if not perfect. Thanks to the resource provided by the BnF, we were able to considerably increase the number of sonnets in the database, with relatively few errors according to our estimation.

As a result, the Oupoco Database is a collection of 4,872 French sonnets. 760 different authors are represented: 4,414 sonnets written by men (660 different authors), and 439 sonnets written by women (107 authors), which leaves 19 sonnets to which we have not been able to assign an author. A specific license related to the source they come from is attached to all the sonnets, but all are freely available and can be reused for free (under the Creative Commons Attribution 4.0 International license), see Table 1.

Source	Number of Sonnets	License/Comments
BnF	3,979	CC-BY-SA-NC
Wikisource	772	CC BY-SA 3.0
Web	67	Source (Blog) cited in the database
Books (anthology)	37	Manual collections of sonnets from different anthologies
Malherbe project	7	No explicit license ( <a href="https://git.unicaen.fr/malherbe/corpus">https://git.unicaen.fr/malherbe/corpus</a> )

Table 1. Sources of the sonnets in the corpus

This database can be used for various teaching and research purposes, especially in the following domains: literature studies, corpus linguistics, digital humanities, arts and technology. All major French poets from the 19th century are included in the database, but also some lesser known ones. Each sonnet is encoded in XML format along with the related metadata; a TEI version of the database is publicly available (for further explanation and access to the source files, see our data paper [Mélodie-Becquet et al 2022] and the related Zenodo repository: <https://zenodo.org/deposit/5646940>).

### 4 Rhyme Analysis

The OuPoCo project has nothing to do with the recent neural approach to poetry generation [Ghazvininejad et al 2017]

[Van de Cruys 2020], but it does require accessing a formal representation of rhymes (as proposed by Beaudouin in 2002<sup>[4]</sup>). We will not detail here the (complex) rules of French versification, but two main difficulties need to be mentioned. First, as is widely known, pronunciation and spelling differ greatly in French, as in English. There is thus a need to first transform written texts into a phonetic transcription. Second, French has a specific and quite complex versification system, which is different (and far more complex) than simple phonetics. The second stage consists then in transforming the phonetic transcription into a representation that corresponds to French versification rules.

In order to do this, the first step is to produce a phonetic transcription of the last word of each line. Phonetization (the process of transforming a word into a phonetic transcription) was done with eSpeak, a free software available on the Web (<http://espeak.sourceforge.net>), that provided satisfactory results on our data. We analyze the whole verse and not only the last word of each line, so that a full rhythmic analysis is possible. However, the phonetic transcription provided by eSpeak is not enough.

23

It is not enough because not everything that has the same sound in French can rhyme: this is, in a way, the beauty of poetry, and specific rules have to be defined to go from a phonetic transcription to a representation that corresponds to the rules of versification. Let's take an example. The two words *aimé* and *aimée* have the same phonetic transcription, but do not rhyme, according to French versification rules (masculine and feminine words, in this case words that end with *-é*, as opposed to *-ée*, do not rhyme); there are also cases where the phonetic transcription is slightly different but the words actually rhyme (for example words ending with the sounds [e] and [ɛ]). All these cases are relatively frequent and must be handled appropriately.

24

A series of rules written in Python had thus to be defined to obtain a proper analysis of rhyme derived from the phonetic transcription (especially of the last word of each line). These rules were manually defined and maintained, as there is no way they could be learnt directly from the data. Some of the rules can be easily derived from a treatise on French versification, but others are directly linked to the output of eSpeak, as we have to overwrite some phonetic distinctions produced by this software that are not relevant to analyze poetry.

25

The sonnet generator uses this analysis to produce sonnets, with different possible structures, respecting the rules of French versification (the code and the resources used, especially the sonnet database, are open source and freely available for research, see: <https://github.com/clement-plancq/oupoco-api>).

26

## 5 Constraints on the Generation Process

In the footsteps of Oulipo and following the comments of Queneau who did not like the idea of purely random poetry generation, we chose to implement constraints to enable the reader to interact with the database, control the generation process in different ways, and discover 19th century French literature from a different and more playful angle.

27

The first constraint corresponds to the different forms of sonnets that have been proposed in the course of history. Giacomo da Lentini, Petrarch, Marot, Peletier, Shakespeare and Spenser all proposed and initiated slightly different rhyming schemes (for example, Marot proposed the following structure: ABBA ABBA CCD EED, while Petrarch proposed several slightly different structures: ABBA ABBA CDE CDE / ABBA ABBA CDC DCD / ABBA ABBA CDE DCE, etc.). All these forms of sonnets are available and the user can choose any structure s/he prefers for generation.

28

The second constraint enables the reader to generate random sonnets from texts within a chosen time framework. In other words, if the reader selects a period of twenty or thirty years that s/he wants to explore on the timeline, sonnets will be generated according to this time frame. This option makes it possible to have a quick overview of the productions of French sonnets over a period. Thanks to this option, it is for example possible to note that very few French sonnets were written at the beginning of the 19th century while their production tended to increase 20 years or so later.

29

The third constraint deals with the authors themselves. The reader can select one or several poets as the source corpus to generate new sonnets. It is thus possible to explore the poems written by Baudelaire for instance, or to combine them with those written by Rimbaud or Verlaine.

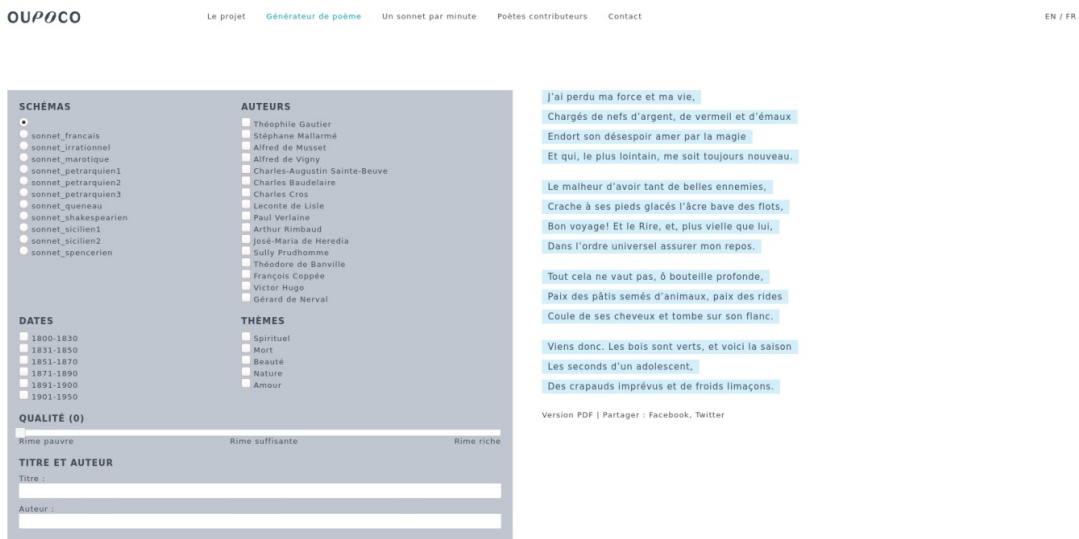
30



The fourth constraint is based on a semantic analysis of the original sonnets. Six main topics were identified (beauty, love, death, melancholy, nature, spirituality), later on reduced to five, as melancholy and death were hard to distinguish (beauty and love are also very close from a linguistic point of view, as well as death and spirituality, but they are nevertheless different enough to be annotated and recognized with sufficient accuracy). A sample of sonnets was annotated using these categories (50 sonnets per category) and a classifier was subsequently trained on the resulting manually annotated corpus. It is thus possible to generate sonnets on a specific topic. Note that the annotation operates at the sonnet level, whereas generation operates at the level of the individual line of verse. However, we assume that the theme gives a general flavor to the text, and that not every line has to be relevant from a thematic point of view

This annotation is however highly subjective. Sometimes most of the poem is on a specific topic, while the last line gives a completely different perspective. See for example *Le dormeur du val*, by Rimbaud. This poem is highly bucolic, clearly about nature since it is about a man peacefully sleeping in a field — except that the last line completely changes the whole perspective, by revealing that the man is in fact a dead soldier. In cases like this, we chose to annotate the dominant topic (the one which covers most of the poem), “nature” in this case (but “death” would also have been relevant, and maybe more relevant if we take into account the deep meaning of the sonnet). All this reveals the subjectivity of any annotation. During the experiment, inter-annotator agreement was not high, but still acceptable and far beyond chance.

In the end, all these constraints are available through a very simple interface on the project web site. Figure 5 gives an overview of this interface. The more constraints one chooses, the fewer the potential sonnets generated (basically, each constraint restricts the possible choices and therefore the number of lines available for generation).



**Figure 5.** The project's web site interface (accessible at: <https://oupoco.org/fr>). On the right, a sonnet produced by the system; on the left, the panel with the different constraints that the user can manipulate.

From a computational point of view, the approach is quite simple. However, it is difficult to control on the fly the number of sonnets that can be generated, depending on the number of constraints chosen by the end user. It is however important to keep track of this, otherwise the end user may frequently arrive at a dead end, with a number of constraints that prevents the possibility of generating new sonnets (a basic rule being that a line cannot rhyme with itself, and cannot be selected twice).

Lastly, we wanted to take gender into account, and highlight female authors, who are underrepresented in most poetry anthologies. A special effort was made to identify sonnets by female authors, as the initial corpus (from Wikisource) contained mostly male authors. Finally, thanks to the data extracted from the BnF corpus and to intensive manual work, we were able to identify 439 sonnets written by 107 female authors (see section 3). This was enough to produce a specific system highlighting the creativity of the female authors, most of whom had remained in the dark so far. The

system is available on the Oupoco web site, in a specific tab<sup>[5]</sup>, see Figure 6.



Figure 6. Female authors in Oupoco

## 6 Evaluation

Poetry is, by nature, hard to evaluate. We did not perform a formal evaluation, as we were unable to provide meaningful evaluation criteria [Gervas 2013]. In our case, the first thing to be done was simply to examine some poems produced by the system, and assess their quality. Due to the nature of the generative process, the generated poems make more or less sense, and are more or less pleasant to read. 36

We identified different causes that affect the quality of the text produced, mainly related to syntax. The most frequent issues are: a pronoun that cannot be linked to any other element in the sentence, or a segment that does not fit with the preceding part of the sentence. Both situations reveal in fact the same problem: one cannot obtain a coherent text simply by assembling smaller chunks of text obtained randomly. This was already one of the main problems of statistical machine translation (based on segments of text that the system had to assemble to form full sentences, see Poibeau (2017): because of the limitations of the approach, the resulting sentences were often slightly incorrect and grammatically odd. This has been largely solved with the modern neural approach, since it is now possible to encode full sentences directly (through vectors and deep learning approaches). 37

But let's come back to our system and examine a specific example. The following poem was fully automatically produced by our system (emphasis by **color difference** added by the authors). 38

Les vendangeurs lassés ayant rompu leurs lignes,  
La trirème d'argent blanchit le fleuve noir  
Sachant que notre cœur s'emporte et se résigne,  
**Étranger tes sanglots** dans le fatal mouchoir.

C'est par un ciel pareil, tout blanc du vol des cygnes,  
**L'éclair laisse, en fuyant,** l'horizon triste et noir.  
Je hais le mouvement qui déplace les lignes,  
Et son verger fleuri n'est qu'un vaste encensoir.

Mais, le corps étendu, n'oublions pas que l'âme,

**Ils** me rendent aveugle au jour qui te proclame  
Si proche et déjà loin de celui que j'aimais.

Haute théologie et solide morale,  
Et maintenant j'habite, hélas! et pour jamais,  
**S'envole, tinte et meurt** dans le ciel rose et pâle.

Any native reader of French will be disturbed by things such as *ils* (*they*) in the third tercet, that cannot be related to anything in the surrounding text. In the first quatrain, *étranger* would fit as an apostrophe, but the rest of the line (...*tes sanglots*...) show that the noun cannot be interpreted as an apostrophe and, in fact, the whole line cannot be linked to what precedes. The same situation arises in the second quatrain (where *l'éclair laisse, en fuyant*... cannot be attached to the surrounding text), and again in the last tercet (*s'envole, tinte et meurt* needs a subject that cannot be found in the text).

39

All these are clear examples that violate syntax and make the whole poem odd and clearly not the result of human inspiration. It is a pity because otherwise, from a semantic point of view, there are some signs of coherence in the fact that the same topics can be found in the different lines. For example, in the second quatrain, *le ciel* (*the sky*), *l'horizon* (*the horizon*), *les lignes* (*the lines*) all refer to the vast surroundings of the poet, whereas *vol des cygnes* (*flight of swans*), *l'éclair* (*lightning*) and *le verger* (*the orchard*) are clear allusions to nature. Even in the first quatrain, a link could be established between *les sanglots* (*sobs*), *le mouchoir* (*the handkerchief*), and *le fleuve noir* (*the black river*), and *notre cœur [qui] s'emporte et se résigne* (*our heart [that] gets carried away and is resigned*), all this being of course signs of sadness.

40

We did not try to solve these issues, for example the syntactic incoherencies. Solving them would considerably complexify the system, which we wanted to keep as simple as possible. Note also that the more constraints there are, the more the number of available lines at each position is reduced. Lastly, working on these issues would involve implementing many heuristics without any clear scientific benefit (except that of producing better poems, which one could say is the main goal, though).

41

The most obvious approach would be to move to neural poetry generation (a completely different approach), which we have explored but decided not to include in this system for the reasons given above: it would complexify the whole architecture and, for example, would require a GPU for training and for testing (as well as for the system in production). We considered that it was more interesting and more promising to explore neural poetry generation separately, as a new system and a new domain of research. We will not address this here as it is beyond the scope of this paper.

42

## 7 Collaboration with Artists

From the beginning, Oupoco was not just an engineering or a research project. Because the project deals with poetry, and because its main objective is to interest people in it, it was important to go beyond a pure algorithm, and also to offer new ways to access poetry.

43

Queneau designed his book as an object, as a device that the reader can manipulate, with which the reader can interact and produce her/his own poetry. It was thus important to us to also materialize the project through an innovative device that would be both intriguing and appealing. In order to do this, we collaborated with two artists who belong to a collective called Atelier Raffard-Roussel. This is what they say themselves about this project and this collaboration:<sup>[6]</sup>

44

We intervened in [the Oupoco Project] to propose a work that could give this research a new form of visibility and materiality.

To do this, we made a simple device that allows a user to generate a poem automatically by turning a crank. Via a small peripheral screen, the user is also informed of the amount of electrical energy that was required to generate this new poem. If they wish, they can try to guess how the *Poetry Box* works through the transparent walls.

This device, which is both mechanical and electronic, is a direct reference to Hans Haacke's

*Condensation Cube*, which can be considered a milestone work of American conceptual art, and which is almost contemporary with Raymond Queneau's book, since it was made in 1965. It seems to us that there is a great similarity between these two works in that each of them reacts, so to speak, to the conditions of their exhibition; the *Condensation Cube* produces water vapour according to the humidity of the place in which it is exhibited and Queneau's book produces poems according to the luck and mood of the reader.

This project was an opportunity for us to tackle the creation of an object that could be both a machine and a work of art and to reflect on the ecological issues involved in such a practice. What is machine-like in works of art and what is sculpture-like in the machines that surround us? How can we make the environmental impact of building a machine or a work of art more visible?

The resulting device, the *Boîte à poésie* (*Poetry box*), is pictured on Figure 7.

45



**Figure 7.** The *Boîte à poésie*, integrating the Oupoco sonnet generator, developed by Atelier Raffard-Roussel. See <http://www.raffard-roussel.com/fr/projets-boite-a-poesie/> for details.

The *Poetry Box* is a very relevant tool to perform public demonstrations, as it is both portable and appealing. The object is also intriguing, since it is transparent and reveals the cables and computer cards (all Raspberry Pi components) it is made of. This proved to be a very efficient tool to engage the public with poetry.

46

In addition to this portable device, we also developed a Twitter bot that posts a quatrain every 6 hours (Figure 8).

47



Figure 8. The Oupoco bot on Twitter

We also released, in collaboration with Riva Illustration a short movie, which explains the project, the approach and the *Boîte à poésie*<sup>[7]</sup>. This short movie is an essential element in the dissemination of our work. It can be consulted online, at any time, without restrictions. It is available in English and in French. This short movie was awarded the prize for “the most entertaining video” during the IJCAI 2023 conference in Macao<sup>[8]</sup> (IJCAI, the 32nd International Joint Conference on Artificial Intelligence, is one of the most prestigious international conferences in Artificial intelligence).

48

These different devices (web site, short movie, *Boîte à poésie*, Twitter bot, etc.) make it possible to propose different kinds of demonstrations, to different audiences, in different settings. They also proved useful to make the project more popular.

49

## 8 Demonstrations and engagement with the public

At the crossroads between surrealism and absurdism, generated sonnets are generally quite funny and convey a dreamlike atmosphere. The themes specific to the Romantic period – such as love, whether it is magnified or lost, death or the fleetingness of time to name but a few – contribute to creating bizarre but intriguing poems. While most of them lack coherence, particularly with respect to the use of punctuation or pronouns, the syntactic confusions this creates can actually reinforce their poetic overtones. The Oupoco system is intended to be presented in front of an audience, to elicit reactions, as a means to make people rediscover poetry and literature.

50

### 8.1 Demonstrations

We gave several demonstrations of the Oupoco project, to different audiences. A number of these presentations occurred online, because of the Covid pandemic. Oupoco was mostly developed between 2018 and early 2020, so became available just at the beginning of the pandemic, which greatly impacted the number of demonstrations we were able to do. However, we seized the opportunities to present the project to researchers and librarians (for example thanks to meetings organized by the French national library, especially during the “Content Generative Models and GLAM” workshop organized by the AI4LAM network, on “artificial intelligence in, for and by libraries, archives and museums”, on 19/04/2022) and also to a wider audience through the participation in open days (at the Ecole normale supérieure and the University Paris Science et Lettres) and science festivals. Demonstrations for researchers were made using traditional presentations on a computer, but the *Boîte à poésie* was the main support for open days and science festivals.

51

## 8.2 Reactions

The main goal of the project, beyond the scientific challenge of automatically analyzing rhymes and detecting the main topic of a poem, was to interest people in poetry and literature in a non-academic way. Reactions were generally very positive and enthusiastic. We did not perform a comprehensive study, but we discuss here some of the observations we made during these demonstrations.

52

Because the machine produces structurally impeccable sonnets, the experiencer is unconsciously encouraged to find coherence in them, simply because we are used to coherence in our everyday life and because incoherence is bewildering [Reinhart 1980]. The second observation we made (which follows on from the previous one to some extent), is a frequent need for the experiencer to go back to the original poem, to see where a given line comes from (on the Oupoco website, “tooltips” always allows the experiencer to go back to the original sonnet). This is encouraging as our main goal is to re-engage people with poetry (showing that poetry is interesting, and surprising).

53

Other reactions were more surprising. Some children were fascinated by the *Boîte à poésie* and the cables inside the object. The fact that computers are made of electronic cards and cables is obvious to us, but computers remain abstract objects for children who have never seen “inside the box”. This experience was a way to explain to them that a computer manipulates information through electricity and that data and programs have to be physically stored somewhere, inside the machine.

54

Last but not least, although we were not able to do many demonstrations in schools due to the pandemic, several people told us that the *Boîte à poésie*, beyond engaging pupils and students with poetry, was also a very good way to raise their awareness of complex linguistic issues such as text coherence. The program frequently produces non optimal texts (with pronouns not referring to any antecedents or with erratic syntactic constructions, as shown in Section 6). Since these texts are produced by a machine, there is no harm in using them to explain why they are problematic, and how they could be improved. The whole process also helps to show that a computer is not perfect in itself: it is just the result of what people (especially, programme designers) do with it.

55

## 8.3 Objections

Even if most of the reactions about the website or about the *Boîte à poésie* were positive, we also had to face a few reservations, generally from teachers and professors of French literature. Their worries concerned the image all this gives of literature, and more specifically poetry. By assembling lines of verse selected from different poems (either randomly or with the light supervision of the reader) one may often produce meaningless texts. This could lead the reader to think that poetry can be anything, that even meaningless text or just a bunch of random words put together can be poetry. In brief, people could lose the sense of what real poetry is.

56

While this has to be taken into account, our experience showed that people are well aware of what we propose through this system. Everybody immediately understood that the sonnets were generated automatically and that they do not have to make sense. In other words, they are not real poems and cannot be confused with poems written by humans. Conversely, as explained in the previous sections, people were often puzzled and wanted to know the source of a given line, and wanted to have a look at the original poem. So, far from being a distraction, the system was a way to reconcile

57

people with poetry, although on a very limited and modest scale. We do not think our system will fundamentally change how people see poetry (a genre with a very limited number of readers nowadays, unfortunately), but it could at least arouse curiosity for this literary genre.

In our opinion, the project is thus not a sacrilege with respect to venerated texts, but a way to make people experience and rediscover poetry.

58

## 9 Discussion

Now that we have described the project in detail, it may be interesting to come back to the rationale behind this project. What have we explored, what have we learnt in the end, beyond developing a piece of software and some devices inspired by Queneau? How should we apprehend the poems created by the machine?

59

As said in the beginning, one piece of inspiration was Kenneth Goldsmith's book *Uncreative writing* (2011). Goldsmith contends that the vast reservoir of pre-existing texts available in digital form allows people (especially writers) to engage with language in novel ways, emphasizing the value of recontextualization and reinterpretation over the creation of entirely new content. This approach challenges conventional ideas about what constitutes originality and innovation in the realm of writing and highlights the evolving relationship between humans and text. Goldsmith provocatively advocates for a shift towards a more "uncreative" approach to writing. He posits that human creativity can also be found in the act of appropriating and remixing existing texts, blurring the lines between authorship and readership. We partly wanted explore this idea through the Oupoco project.

60

This may also remind the reader of Charles O. Hartman's book, *Virtual Muse: Experiments in Computer Poetry* (1996). In this book, Hartman explores technology and literature, showcasing the transformative potential of computers in redefining the very essence of poetry. Through meticulous analysis and experimentation, he demonstrates how algorithms and computational processes can be harnessed to generate innovative and unexpected poetic forms, challenging established norms and expanding the horizons of creative expression. More specifically, his first experiment was in a way a precursor of our own: a program named "RanLines" was designed to store a collection of 20 lines and subsequently to fetch one of these lines at random whenever the user pressed a key, so as to produce a kind of poem. Through this program, the author explores a "simple sort of randomness" that reminds him of the Greek oracles.

61

What is even more interesting are Hartman's reflections on the results he got. He observed that "It's hard to put together two words that don't make some kind of sense to the willing reader" [Hartman 1996, 31]. He also adds that: "The more discrete and self-contained the syntax of the line (complete clause, complete prepositional phrase), the more easily it joins with lines before and after", which is also what we observed, and one of the limits of our experiments (when a linguistic element refers to other ones outside the line, coherence – or simple linguistic well formedness – is often lacking). But beyond this issue, we agree with Hartman that these experiments open new perspectives to the willing reader. As Hartman suggests, the reader's imagination is challenged, as it is nearly always possible to find some meaning, draw some connexions and interpret combined verses. At the same time, our experiments showed that the application is also a strong incentive to rediscover poetry: very often, people are intrigued by a line and want to go back to the original poem.

62

To conclude, we can say we have explored in a new manner the idea of potential literature, by producing new sonnets from a very well-defined set of constraints, following in this sense the program put forward by Queneau more than 60 years ago.

63

## 10 Conclusion

The main interest of the Oupoco project is to present French poetry through a new and original setting. With our system, poetry is not just a literary genre, but a dynamic object that can be manipulated and experienced. For many people, poetry is seen at best as something related to school education, at worst as something boring and uninteresting from the past. Our new setting, in itself, makes it possible to show that playing with poetry can be fun. Our setting puts the notion of text coherence in perspective since the result of the generator can be more or less satisfactory from a

64

semantic point of view [Reinhart 1980]. In the future, we plan to study the potential impact of our system in different (real world) contexts, especially in educational settings.

## Acknowledgements

This work received support from EUR Translitterae (Ecole universitaire de recherche, program “Investissements d’avenir” ANR-10-IDEX-0001-02 PSL\* and ANR-17-EURE-0025) and from the CNRS through the International Research Network Cyclades (Corpora and Computational Linguistics for Digital Humanities). This work was also supported in part by the French government under the management of the Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR19-P3IA-0001 (PRAIRIE 3IA Institute).

65

## Author Roles

Frédérique Mélanie-Becquet: conceptualization, data curation and supervision

Clément Plancq: conceptualization and software development

Claude Grunspan: data curation

Mylène Maignant: conceptualization and data curation

Matthieu Raffard: conception and development of the *Poetry Box (Boîte à poésie)*

Mathilde Roussel: conception and development of the *Poetry Box (Boîte à poésie)*

Fiammetta Ghedini: conception and development of the Oupoco video clip

Thierry Poibeau: conceptualization, funding acquisition, supervision and writing

## Notes

[1] Oulipo (“Ouvroir de littérature potentielle”; roughly translated: “workshop of potential literature” and stylized Oulipo) is a loose gathering of (mainly) French-speaking writers and mathematicians who seek to create works using constrained writing techniques. (Wikipedia)

[2] The decision of the court can be found online: <https://www.legalis.net/jurisprudences/tribunal-de-grande-instance-de-paris-ordonnance-de-refere-du-5-mai-1997/>

[3] In the original French: “la recherche de formes, [...] de structures nouvelles et qui, ensuite, pourront être utilisées par les écrivains de la façon qui leur plaira”

[4] For more details, see The Metrometer, a tool developed to analyze French verse [Beaudouin and Yvon 1996] that is however, obsolete today.

[5] <https://oupoco.org/fr/sonnet-feminin/index.html> (last consulted, 7/11/2022).

[6] *Nous sommes intervenus dans le cadre [du projet oupoco] pour proposer une œuvre qui puisse donner à cette recherche une nouvelle forme de visibilité et de matérialité. Pour cela, nous avons réalisé un dispositif simple qui permet à un utilisateur de générer un poème automatiquement en tournant une manivelle. Par l’intermédiaire d’un petit écran périphérique, l’utilisateur est également informé de la quantité d’énergie électrique qu’il a été nécessaire de mobiliser pour générer ce nouveau poème. S’il le souhaite, il peut essayer de deviner le fonctionnement de la Boîte à Poésie à travers les parois transparentes. Ce dispositif, à la fois mécanique et électronique, fait directement référence au Condensation Cube de Hans Haacke qui peut être considéré comme une œuvre jalon de l’art conceptuel américain, et qui est presque contemporain du livre de Raymond Queneau puisqu’elle a été réalisée en 1965. Il nous semble que l’on peut voir une grande similitude entre ces deux œuvres dans la mesure où chacune d’entre elles réagit pour ainsi dire aux conditions de leur exposition ; le Condensation Cube produit de la vapeur d’eau en fonction du taux d’hygrométrie du lieu dans lequel il est exposé et le livre de Queneau produit des poèmes en fonction de la chance et l’humeur du lecteur. Ce projet a été pour nous l’occasion de nous confronter à la réalisation d’un objet qui puisse être à la fois une machine et une œuvre d’art et de réfléchir aux enjeux écologiques qu’implique une telle pratique. Qu’est-ce qui relève de la machine dans les œuvres d’art et en retour qu’est-ce qui relève de la sculpture dans les machines qui nous entourent ? Comment rendre mieux visible l’impact environnemental qu’entraîne la construction d’une machine ou d’une œuvre d’art?*

[7] <https://riva-illustrations.com/animations-oupoco/> (last consulted 7/11/2022).



## Works Cited

- Baillehache 2021** Baillehache, J. (2021) "The Digital Reception of A Hundred Thousand Billion Poems". *Sens public*, pp. 1–13. Available at: <https://doi.org/10.7202/1089666ar>.
- Beaudouin 2002** Beaudouin, V. (2002) *Mètre et rythmes du vers classique. Corneille et Racine*. Paris : Honoré Champion (Lettres numériques).
- Beaudouin and Yvon 1996** Beaudouin, V. and Yvon F. (1996) "The Metrometer: a Tool for Analysing French Verse" *Literary and Linguistic Computing*, 11(1), pp. 23–31. Available at: <https://doi.org/10.1093/lilc/11.1.23>.
- Bens 2005** Bens, J. (2005) *Genèse de l'Oulipo: 1960-1963. Circulaire du 28 août 1961*. Le Pré-Saint-Gervais: Le Castor Astral.
- Berkman 2017** Berkman, N. (2017) "Digital Oulipo: Programming Potential Literature", *Digital Humanities Quarterly*, 11(3). Available at: <http://www.digitalhumanities.org/dhq/vol/11/3/000325/000325.html> (Accessed 27 August 2023).
- Berkman 2022** Berkman, N. (2022) *Oulipo and the Mathematics of Literature*. Bern: Peter Lang.
- Bloomfield and Campaignolle 2016** Bloomfield, C., and Campaignolle, H. (2016) "Machines littéraires, machines numériques : l'Oulipo et l'informatique" in C. Reggiani and A. Schnaffner (eds.) *Oulipo mode d'emploi*. Paris: Honoré Champion, pp. 319–36.
- Charbonnier 1962** Charbonnier, G. (1962) *Entretiens avec Raymond Queneau*. Paris : Gallimard.
- Denize and Magné 1999** Denize, A. and Magné, B. (1999) *Machine à écrire* [CD-Rom]. Paris: Gallimard.
- Derrida and Ronell 1980** Derrida, J. and Ronell, A. (1980) "On narrative: The law of genre", *Critical Inquiry*, 7(1), pp. 55–81.
- Fournel 1972** Fournel, P. (1972) *Clés pour la littérature potentielle*. Paris: Denoël.
- Gervas 2013** Gervas, P. (2013) "Computational modelling of poetry generation" in *Artificial Intelligence and Poetry Symposium*, AISB Convention, University of Exeter.
- Ghazvininejad et al 2017** Ghazvininejad M., Shi X., Priyadarshi J., and Knight K. (2017). "Hafez: an interactive poetry generation system" in *Proceedings of ACL 2017, System Demonstrations*. Vancouver: Canada, pp. 43–48.
- Goldsmith 2011** Goldsmith, K. (2011) *Uncreative Writing : Managing Language in the Digital Age*. Columbia University Press: New York.
- Goldsmith 2019** Goldsmith, K. (2019) *What Big Data Can't Do: The Pleasure of the (Perverse) Text*. Available at: <https://savoirs.ens.fr/expose.php?id=3659> (Accessed 27 August 2023).
- Hartman 1996** Hartman C.O. (1996) *Virtual Muse: Experiments in Computer Poetry*. Middletown: Wesleyan University Press.
- Leonardo/Olats and Bootz 2006** Leonardo/Olats and Bootz, P. (2006) "Qu'est-ce que la littérature générative combinatoire?" In P. Bootz, *Les Basiques : la littérature numérique*. Publication en ligne : [http://archive.olats.org/livresetudes/basiques/litteraturenumerique/10\\_basiquesLN.php](http://archive.olats.org/livresetudes/basiques/litteraturenumerique/10_basiquesLN.php) (consulté le 3/11/2022).
- Mélanie-Becquet et al 2022** Mélanie-Becquet, F., Grunspan, C., Maignant, M., Plancq, C., and Poibeau, T. (2022) "The Oupoco Database of French Sonnets from the 19th Century", *Journal of Open Humanities Data*, 8(0), pp. 1–5. Available at: <https://doi.org/10.5334/johd.89>.
- Oulipo 1973** Oulipo. (1973) *La littérature potentielle (Créations Re-créations Récréations)*. Paris: Gallimard.
- Oulipo 1981** Oulipo. (1981) *Atlas de littérature potentielle*. Paris: Gallimard.
- Oulipo 2002** Oulipo. (2002) *Abrégé de littérature potentielle*. Paris: Éditions Mille et une nuits, n° 379.
- Poibeau 2017** Poibeau, T. (2017) *Machine Translation*. Cambridge: MIT Press.
- Queneau 1961** Queneau, R. (1961) *Cent mille milliards de poèmes*. Paris: Gallimard.
- Reinhart 1980** Reinhart, T. (1980) "Conditions for text coherence". *Poetics Today*, 1(4), pp. 161–180.

**Starynkevitch 1990** Starynkevitch, D. (1990) "The SEA CAB 500 Computer", *IEEE Annals of the History of Computing*, 12(1), pp. 23-29. Available at: 10.1109/MAHC.1990.10008.

**Van de Cruys 2020** Van de Cruys, T. (2020) "Automatic poetry generation from prosaic text" in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Available at: <https://aclanthology.org/2020.acl-main.223/> (Accessed 10 October 2023).



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.