

Les méthodes de génération de texte

Le cut-up

Le cut-up est une technique littéraire, inventée par l'auteur et artiste Brion Gysin, et expérimentée par l'écrivain américain William S. Burroughs, où un texte original se trouve découpé en fragments aléatoires puis ceux-ci sont réarrangés pour produire un texte nouveau.

Le cut-up

fonctionnement

- 1) on sépare la phrase en mots ou en bout de phrases
- 2) on mélange et on réordonne pour créer un nouveau texte

Exemple: Salut c'est moi



["salut", "c'est", "moi"]

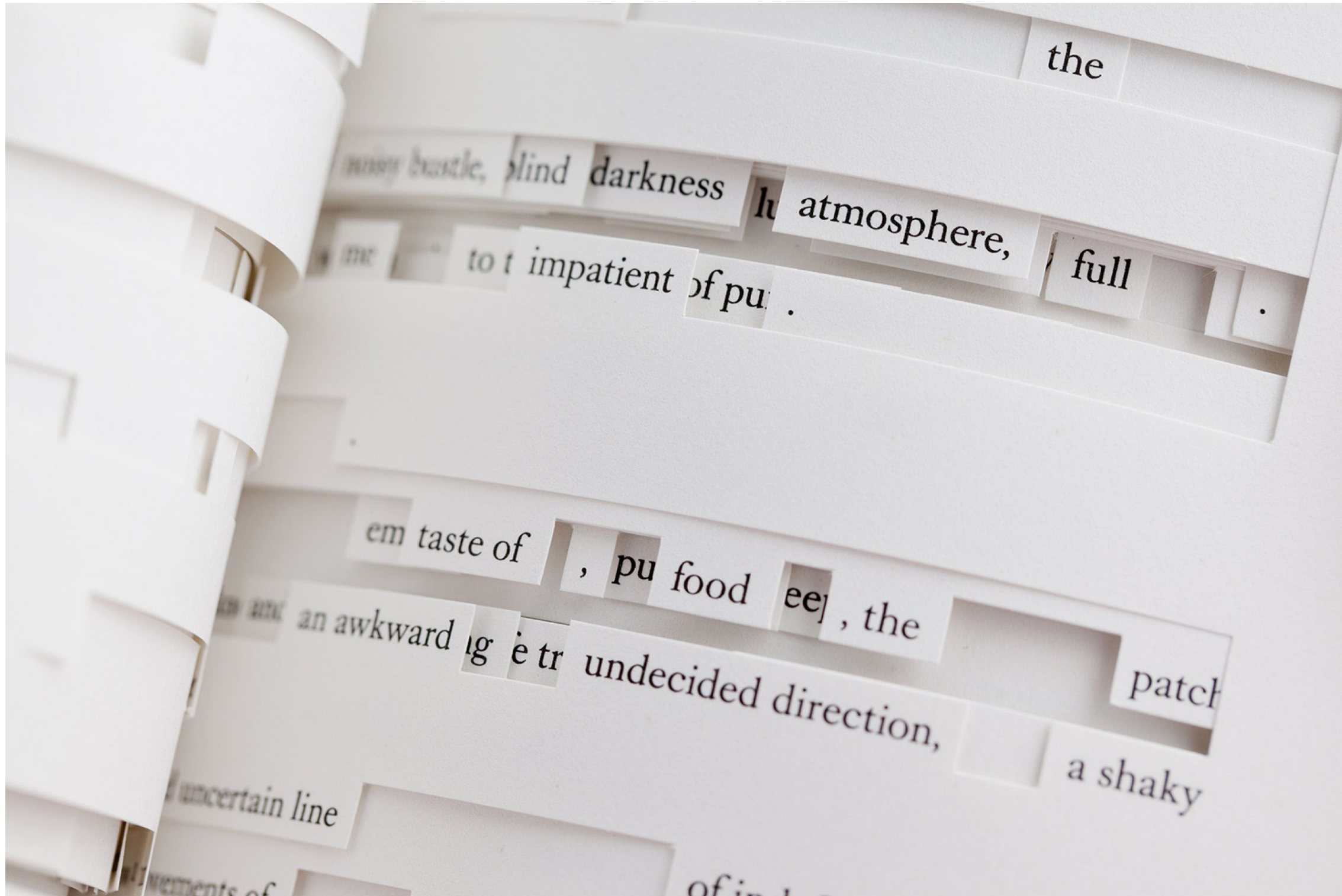


moi c'est salut

L'effacement

L'effacement est une forme de poésie fabriquée en effaçant des mots d'un texte existant en prose ou en vers et en encadrant le résultat sur la page comme un poème. Les résultats peuvent rester tel quel ou ils peuvent être arrangés en lignes et/ou en strophes.

L'effacement



Jonathan Safran Froer, *Tree of Code*

L'effacement

fonctionnement

- 1) on sépare la phrase en mots ou en bout de phrases
- 2) on supprime certains mots ou certains bout de phrases

Exemple: Salut c'est moi

↓
["salut", "moi"]

↓
salut moi

Exemple: Salut c'est moi

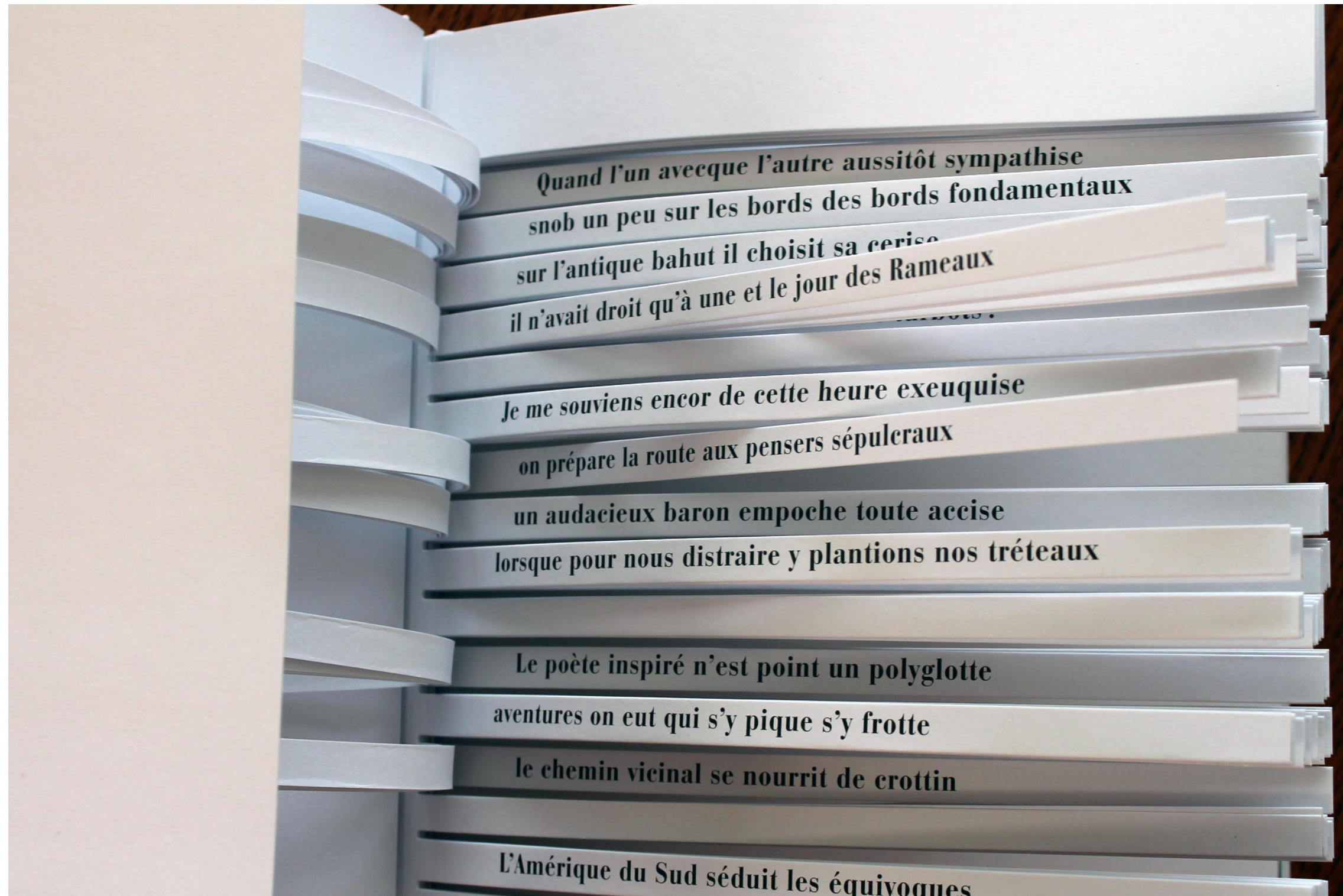
↓
["salut", ~~"c'est"~~, "moi"]

↓
salut moi

La combinaison

La combinaison consiste à combiner des phrases, des bouts de phrases ou des mots afin de construire un nouveau texte.

La combinaison



Raymond Queneau, *Cent mille milliards de poèmes*

La combinaison

fonctionnement

- 1) on définit une liste de phrases à combiner
- 2) on mélange aléatoirement l'ordre de ces phrases
- 3) on met les phrases bout à bout

Exemple: ["Demain c'est loin.", "Peut-être qu'il pleuvra.", "J'irai à la piscine.", "Quoi qu'il arrive."]



J'irai à la piscine. Peut-être qu'il pleuvra. Quoi qu'il arrive. Demain c'est loin.

Chaînes de Markov

Les chaînes de Markov sont un principe mathématiques, elles permettent la prédiction d'un état futur basé sur les caractéristiques d'un état présent. Les chaînes analysent les mots et leurs probabilités d'apparition les uns à la suite des autres. Elles génèrent ensuite des chaînes de mots qui sont probablement liés.

La génération est entièrement aléatoire et basée sur les probabilités d'associations entre chaque mot.