Document généré le 7 juil. 2025 11:52

Communitas
Théories et pratiques de la normativité

Le Schtroumpf civil
The “Schtroumpf civil”

Guillaume Simiand

Volume 4, numéro 1, 2023

Contraintes et droits

URI : https://id.erudit.org/iderudit/1108315ar
DOI : https://doi.org/10.7202/1108315ar

Aller au sommaire du numéro

Éditeur(s)

UQÀM Département des sciences juridiques
UQÀM Faculté de science politique et de droit

ISSN

2563-9250 (numérique)

Découvrir la revue

Citer ce document

Simiand, G. (2023). Le Schtroumpf civil. *Communitas*, *4*(1), 4–15. https://doi.org/10.7202/1108315ar

Résumé de l'article

Le Schtroumpf civil est la première traduction intégrale des Codes civils français et québécois en langue schtroumpf. Si cette démarche, clin d’œil burlesque à la centralité historique du Code civil dans la sphère francophone, est sa propre fin, cet article revient sur la réflexion théorique sur les régularités de la langue du droit qu’a exigée cette traduction. Plus que sur la langue schtroumpf originelle inventée par Peyo, elle s’appuie sur sa réinterprétation systématique par Umberto Eco, qui s’est penché à plusieurs reprises au cours de sa carrière sur une langue dont l’ambiguïté permanente devrait obérer toute compréhension, mais qui reste parfaitement intelligible pour les lecteurs. C’est, raison de l’intérêt du linguiste, que la langue schtroumpf illustre parfaitement la place prépondérante de l’activité interprétative du lecteur, et notamment le renvoi hors texte à un encodage commun des références situationnelles et textuelles qui permet la restitution du sens. Par de multiples aspects, ce fonctionnement de la langue schtroumpf rappelle des difficultés similaires soulevées par l’interprétation de la langue du droit, liées à sa clarté, à sa lisibilité et à son intelligibilité. Pour générer automatiquement la traduction des Schtroumpfs civils, une stratégie fondée sur l’apprentissage machine est mise en œuvre ; pour chaque mot susceptible d’être remplacé par schtroumpf, on demande à un modèle de langue une prédiction ; si le texte du Code et la prédiction coïncident, le mot est schtroumpfé. Ce dispositif forme une variante automatisée du test de closure, utilisé depuis les années cinquante comme outil de mesure de la lisibilité et, pour certains auteurs, de l’intelligibilité des textes. Le Schtroumpf civil, quoique ce ne soit pas son objet principal, rendrait ainsi visible le degré de proximité syntactique de la langue des législateurs et de la langue générale, une approche potentiellement nouvelle d’un des versants de l’intelligibilité du texte législatif.

© Guillaume Simiand, 2023

Ce document est protégé par la loi sur le droit d’auteur. L’utilisation des services d’Érudit (y compris la reproduction) est assujettie à sa politique d’utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l’Université de Montréal, l’Université Laval et l’Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Le Schtroumpf civil

Simiand Guillaume[1]

RÉSUMÉ

Le Schtroumpf civil est la première traduction intégrale des Codes civils français et québécois en langue schtroumpf. Si cette démarche, clin d'œil burlesque à la centralité historique du Code civil dans la sphère francophone, est sa propre fin, cet article revient sur la réflexion théorique sur les régularités de la langue du droit qu'a exigée cette traduction. Plus que sur la langue schtroumpf originelle inventée par Peyo, elle s'appuie sur sa réinterprétation systématique par Umberto Eco, qui s'est penché à plusieurs reprises au cours de sa carrière sur une langue dont l'ambiguïté permanente devrait obérer toute compréhension, mais qui reste parfaitement intelligible pour les lecteurs. C'est, raison de l'intérêt du linguiste, que la langue schtroumpf illustre parfaitement la place prépondérante de l'activité interprétative du lecteur, et notamment le renvoi hors texte à un encodage commun des références situationnelles et textuelles qui permet la restitution du sens. Par de multiples aspects, ce fonctionnement de la langue schtroumpf rappelle des difficultés similaires soulevées par l'interprétation de la langue du droit, liées à sa clarté, à sa lisibilité et à son intelligibilité. Pour générer automatiquement la traduction des Schtroumpfs civils, une stratégie fondée sur l'apprentissage machine est mise en œuvre ; pour chaque mot susceptible d'être remplacé par *schtroumpf*, on demande à un modèle de langue une prédiction ; si le texte du Code et la prédiction coïncident, le mot est *schtroumpfé*. Ce dispositif forme une variante automatisée du test de *closure*, utilisé depuis les années cinquante comme outil de mesure de la lisibilité et, pour certains auteurs, de l'intelligibilité des textes. Le Schtroumpf civil, quoique ce ne soit pas son objet principal, rendrait ainsi visible le degré de proximité syntactique de la langue des législateurs et de la langue générale, une approche potentiellement nouvelle d'un des versants de l'intelligibilité du texte législatif.

MOTS-CLÉS : Schtroumpf, légistique, lisibilité, intelligibilité, apprentissage machine.

ABSTRACT

The "Schtroumpf civil" is the first complete translation of the French and Québécois Codes Civils into the Smurf ("Schtroumpf" in French) language. If this endeavour, a burlesque nod to the historical centrality of the Code civil in the French-speaking sphere, is its own end, this article returns to the theoretical reflection on the regularities of the language of law that this translation

required. It is based, more than on the original Smurf language invented by Peyo, on its systematic reinterpretation by Umberto Eco, who came back several times during his career to a language whose permanent ambiguity should hamper any understanding, but which remains perfectly intelligible for its readers. The Smurf language perfectly illustrates the preponderant role of the interpretative activity of the reader in the formation of meaning (the source of its interest to Eco). In multiple aspects, this logic of the Smurf language recalls similar questions raised by the interpretation of the language of law, linked to its clarity, readability, and intelligibility. To automatically generate the translation of the *Schtroumpfs civils*, a strategy based on machine learning is implemented; for each word potentially replaced by *schtroumpf*, a call is made to a language model for prediction; if the text of the Code and the prediction coincide, the word is *smurfed*. This system forms an automated variant of the closure test, used since the 1950s as a tool for measuring the readability and, for some authors, the intelligibility of texts. The *Schtroumpf civil*, although it is not its main goal, thus makes visible the degree of syntactic proximity of the language of legislators and the general language, a potentially novel approach to one of the aspects of the intelligibility of the legislative text.

KEYWORDS: Smurf, legistics, readability, intelligibility, machine learning.

[1] Professeur agrégé à l'Université Paris 1 Panthéon-Sorbonne

Introduction

« Les lois de police et de sûreté obligent tous ceux qui schtoumpfent le territoire », « La schtroumpf ne dispose que pour l'avenir ; elle ne schtroumpfe point de schtroumpf rétroactif » pour les français ; « Toute perschtroumpf est titulaire d'un patrimoine. Celui-ci schtroumpfe faire le schtroumpf d'une division ou d'une affectation, mais dans la seule schtroumpfure préschtroumpfée par la schtroumpf » pour les québécois : même en langue schtroumpf[1], les grandes pages du droit sont immédiatement reconnaissables. Nous présentons ici le Schtroumpf civil, première traduction intégrale des codes civils français et québécois en langue schtroumpf[2].

La question qui s'impose – pourquoi traduire le Code civil en langue schtroumpf ? – appelle plusieurs types de réponses, que l'on déclinera ici ; peut-être apprendrons-nous quelque chose, au passage, sur la langue du droit et son rapport à la langue générale. Conformément à l'habitude française de rédiger des plans en deux parties, nous montrerons d'abord que le Schtroumpf civil ne sert à rien (1), avant d'essayer de prouver qu'il peut servir à quelque chose (2).

1. Le Schtroumpf civil ne sert à rien

Une première réponse à la question « pourquoi ? » pourrait être « pourquoi pas ? » : acte gratuit, fête immotivée. La référence universelle et permanente dans le monde francophone au grand œuvre juridique qu'est le Code civil appelle le recours à l'hyperonyme ultime qu'est le mot *schtroumpf* : tout juriste comprend immédiatement de quoi l'on parle si l'on mentionne le Schtroumpf civil.

1.1. Une avancée vers l'état de droit pour les Schtroumpfs

Si l'on se place du point de vue des petits êtres bleus inventés par Peyo, pseudonyme de Pierre Culliford, il s'agit sans conteste d'un grand pas en direction de l'entrée des Schtroumpfs dans l'État de droit. Depuis 1958 et la sortie des premiers albums, des auteurs divers ont voulu voir dans la microsociété schtroumpf la réalisation concrète ou la métaphore d'idéologies divergentes, voire opposées, inventoriées et analysées en 2016 par Antonio Dominguez Leiva et Sébastien Hubier dans *Schtroumpfologies* (DOMINGUEZ LEIVA & HUBIER, 2016). Loin de nous l'idée de nous mettre à la queue de cette cohorte ; mais force est de constater que la société schtroumpf est une gérontocratie où la séparation des pouvoirs n'existe pas. Le Grand Schtroumpf en détient tous les leviers, comme l'illustre l'album

[1] On emploiera dans cet article les expressions « langue schtroumpf » et « langue du droit » dans la mesure où il s'agit d'expressions reçues, mais il serait plus exact pour l'une comme pour l'autre de parler de « langage », voir le 1.5 du présent article.

[2] Disponibles en texte intégral sur https://github.com/g-simiand/Schtroumpf_civil

Schtroumpf vert et vert schtroumpf. Les Schtroumpfs vivaient heureux sans lois, sous l'autorité charismatique du Grand Schtroumpf, comme diraient Jean-Jacques Rousseau et Max Weber ; ils vivront, à n'en pas douter, encore plus heureux avec la loi, sous le régime légal-rationnel du Schtroumpf civil.

1.2. Un projet oudropien

À un autre horizon, on peut voir le Schtroumpf civil comme un maillon dans une série de projets conçus au sein de l'Oudropo[3], groupe de recherche et de création juridique qui, au moyen d'outils inspirés de la littérature à contraintes pratiquée par l'Oulipo, travaille les textes du droit pour en faire jaillir des facettes nouvelles. Utiliser la langue du droit comme matériau, apporter des solutions non sollicitées à des problèmes imaginaires, dans une logique burlesque mais pas seulement, opérer un décalage radical pour aborder un corpus trop connu sous un angle nouveau, brancher des outils et des approches qui n'étaient pas faits pour travailler ensemble : pas de doute, la démarche est bien oudropienne.

Comme dans beaucoup de projets oudropiens, ce sont les régularités de la langue juridique qui sont au cœur de l'exploration oblique qu'opère le Schtroumpf civil. Vue de loin, la langue schtroumpf est répétitive et obscure : un peu, pour qui la perçoit de l'extérieur, comme la langue du droit[4]. Mais si le Schtroumpf civil est une satire de la langue juridique et de certaines formes de solutionnisme technologique (la traduction est réalisée par un algorithme), ainsi qu'un clin d'œil à l'omniprésence de la référence au Code civil dans le contexte français, peut-être pouvons-nous au-delà du burlesque retirer quelque chose du choc de ces deux langues.

1.3. Fonctionnement de la langue schtroumpf originelle

Il est bon, pour comprendre ce rapprochement, de se pencher sur la réalité de la langue schtroumpf plutôt que sur son halo symbolique. Comment fonctionne la langue schtroumpf, ou, pour le dire autrement, quelles sont les règles qui autorisent la substitution de « schtroumpf » à un mot de la langue ordinaire ? Dans sa version originale, celle de Peyo, elle n'est en réalité que très exceptionnellement obscure. La substitution n'obéit à aucune règle établie : elle relève du libre choix de l'auteur à la recherche d'un effet comique. Dans les albums, on observe que les verbes, à l'exception des auxiliaires, sont très souvent substitués, y compris les infinitifs et les participes passés. Les substantifs, quant à eux, sont *schtroumpfés* essentiellement dans les expressions figées,

[3] https://www.oudropo.com/, consulté le 9/11/2022.

[4] Le constat est fréquemment dressé par les juristes eux-mêmes ; on citera parmi tant d'autres J.-C. Gémar, qui écrivait en 1990 dans « Les fondements du langage du droit comme langue de spécialité » : « Entre toutes les langues de spécialité possibles (médicale, économique, scientifique ou technique, etc.), il est de notoriété publique que celle du droit est une des plus complexes, que les juristes pratiquent un discours souvent obscur et tortueux à souhait, et cela dans la plupart des langues véhiculaires, en Occident tout au moins. » (GÉMAR, 1990, p. 719)

dont les nombreux proverbes qui émaillent le texte. Les vignettes permettent de restituer aisément le sens des verbes et de jouir des effets de connivence qu'engendre ce cryptage/décryptage – c'est le contexte immédiat qui porte le sens[5]. Pour la même raison, les adverbes en *-ment*, adverbes de manière directement reliés à l'action et pour cette raison aisément restituables, sont aussi très souvent « schtroumpfés ». Quant aux proverbes, ils produisent le même effet de connivence et de reconnaissance à l'échelle de la langue plus que du contexte, pour un effet comique à destination sans doute des lecteurs adultes. Les épisodes où le comique provient de l'obscurité de la langue schtroumpf sont rares[6].

On le comprend aisément : l'inintelligibilité, comme procédé comique, ne peut être utilisée qu'avec la plus grande parcimonie ; dans l'exemple cité, la complexité de la langue se veut mimétique de la complexité technique. La langue schtroumpf de Peyo cherche à produire des effets de décryptage et de connivence avec le plus grand nombre. Ce n'est donc pas l'obscurité ni l'inintelligibilité qui caractérisent la langue schtroumpf originelle, mais au contraire la transparence complète qu'elle vise, pour peu qu'on puisse s'appuyer sur le contexte de l'énonciation, contexte référentiel ou linguistique selon les cas.

1.4. La langue schtroumpf d'Umberto Eco

La langue schtroumpf présente donc un intéressant paradoxe : là où les langues naturelles fuient l'ambiguïté, et en premier lieu l'homonymie, la langue schtroumpf les multiplie, et reste parfaitement intelligible. C'est ce caractère paradoxal qui a suscité l'intérêt des linguistes et des sémiologues, au premier rang desquels Umberto Eco. Eco a consacré deux articles assez longs à la langue schtroumpf : le premier, en 1979, était intitulé « Schtroumpf und Drang » (ECO, 1979), clin d'œil au mouvement *Sturm und Drang*, tempête et passion, premier mouvement du romantisme allemand et univers fort différent, au premier abord, de celui des Schtroumpfs. Ce texte, paru dans la revue *Alfabeta*, se présente comme une recension de trois albums récemment publiés ; mais au fil de ses 15 000 signes, Eco parle essentiellement de la langue schtroumpf. Il situe d'emblée son intérêt pour cette « méditation pratique sur le fonctionnement contextuel du langage » dans le contexte de ses travaux sur la réception/interprétation des textes ; son *opus majus*, *Lector in fabula : la cooperazione interpretativa nei testi narrativi*[7], venait d'être publié. La force du

[5] Le moment de naissance mythique du langage schtroumpf l'illustrait déjà, avant même les albums : en 1957, alors que Peyo et Franquin mangent ensemble, Peyo demande la salière à Franquin. Mais, butant sur le mot, il abandonne et dit « Passe-moi le schtroumpf » (DAYEZ, 2003).

[6] On peut noter la planche intitulée « Ce qui se schtroumpfe bien se schtroumpfe clairement » dans l'album *La Soupe aux Schtroumpfs*, où l'on trouve le dialogue suivant : « - Bon ! Eh bien lorsque tu schtroumpfes sur ce schtroumpf, tu schtroumpfes une série de schtroumpfs enschtroumpfés entre eux qui, en se désenschtroumpfant, schtroumpfent par leur schtroumpf de la schtroumpf au schtroumpf ! Tu schtroumpfes ? / - Oui, oui ! / - Il n'a pas l'air d'avoir schtroumpfé !... C'est pourtant clair, non ? » Le clin d'œil à Boileau souligne que l'on se trouve dans une parenthèse métalinguistique (métapoétique ?).

[7] Traduit en français sous le titre *Lector in fabula : Le rôle du lecteur, ou, La coopération interprétative dans les textes narratifs*.

contexte dans la compréhension de la langue schtroumpf s'intégrait parfaitement avec son affirmation de la prépondérance du rôle du lecteur, du récepteur dans la formation du sens.

Eco revisite la question de la langue schtroumpf vingt ans plus tard, dans un texte plus théorique, *Kant et l'ornithorynque* (ECO, 1997) ; il y insiste sur le caractère impossible de la langue schtroumpf : le mot « schtroumpf », une seule syllabe, un son vocalique pour six consonantiques, est du point de vue morphologique exactement ce que les langues romanes préfèrent éviter. En linguiste, Eco va s'efforcer de proposer une grammaire de ce que serait une langue schtroumpf autonome. Abordée pour sa différence avec le langage ordinaire, elle fonctionne autour d'une règle cardinale : remplacer chaque terme par des flexions du mot *schtroumpf* chaque fois que c'est possible sans tomber dans une ambiguïté excessive.

De ce point de vue, comme le note Eco, les Schtroumpfs illustrent parfaitement le rôle du récepteur dans le travail d'interprétation : « Les Schtroumpfs sont fidèles à la recherche de la linguistique textuelle et de la pragmatique du discours, pour lesquelles chaque texte est une machine paresseuse qui nécessite une coopération interprétative active de la part de son destinataire, appelé à relier les fragments de texte à d'autres textes antérieurs et présupposés[8]. » Dans un des albums, écrit Eco, Gargamel, l'ennemi juré des Schtroumpfs, parvient à se faire passer pour l'un d'eux ; il a l'apparence d'un Schtroumpf, mais son incapacité à parler leur langue le trahit. Pourquoi est-il incapable de parler la langue schtroumpf ? Faute, pour le linguiste, de contexte culturel partagé : « Et Gargamel ne sait pas quand il pourra le faire [parler la langue schtroumpf]. Pourquoi ? Car pour pouvoir parler schtroumpf il faut connaître non seulement la grammaire de la langue-base, mais aussi ses règles (hypercodées) d'intertextualité. Il ne connaît pas ces portions de langage déjà parlé qui permettent aux Schtroumpfs de schtroumpfer quand existe comme un fait acquis dans leur fond de compétence une expression standard donnée[9]. »

1.5. Similarités et différences des langues du droit et du schtroumpf

Les notions de clarté, de lisibilité et d'intelligibilité de la langue du droit ont été problématisées depuis longtemps par les jurilinguistes (entre autres

[8] Notre traduction, comme pour les citations suivantes du même texte. La version originale : « I puffi sono fedeli alle ricerche di linguistica testuale e di pragmatica del discorso, per cui ogni testo è una macchina pigra che richiede una attiva cooperazione interpretativa da parte del suo destinatario, chiamato a connettere le porzioni testuali ad altri testi precedenti e presupposti. »

[9] « E Gargamella non sa quando può farlo. Perché? Perché per poter parlare puffo occorre non solo conoscere la grammatica della lingua-base, ma anche le sue regole (ipercodificate) di intertestualità. Egli non conosce quelle porzioni di lingua già parlata che permettono ai puffi di puffare quando sullo sfondo della loro competenza esiste come acquisita una data espressione standard. » Eco précise le rapport entre contexte et intertexte autour de cet exemple dans *Kant et l'ornithorynque* : « Mais le problème de Gargamel, c'est bien évidemment qu'il trouve tous les contextes ambigus, c'est-à-dire incompréhensibles, et ce pour la simple et bonne raison qu'il n'a pas d'informations intertextuelles. »

FLÜCKIGER, 2006, 2015 ; GÉMAR, 1990, 2019) ; la langue schtroumpf n'est pas claire, mais elle est paradoxalement parfaitement intelligible, puisque selon la définition d'Eco, elle n'est cryptée que lorsque son déchiffrement n'est pas problématique.

Pour certains auteurs, langue du droit et langue schtroumpf présenteraient dans certains contextes une circularité similaire : peu de gens ont eu l'idée de croiser le monde des Schtroumpfs et celui du droit, mais c'est ce qu'a fait Pierre Sargos, Président de chambre honoraire à la Cour de cassation, dans un article de 2008 intitulé « La causalité en matière de responsabilité ou le "droit Schtroumpf" » ; il y critique, autour de la notion de cause, une polysémie préjudiciable (« Il y a, aussi bien dans la doctrine que dans la jurisprudence judiciaire et administrative, une sorte de "langage Schtroumpf" en matière de cause. »), même si nous avons noté que c'est seulement par exception que les Schtroumpfs ne se comprennent pas entre eux – contrairement à nous peut-être.

Autre point commun, langue schtroumpf et langue du droit sont des langues parasitaires, dans la mesure où elles se développent à partir du français ordinaire, auquel elles empruntent l'essentiel du vocabulaire et de la syntaxe : elles n'ont pas d'existence autonome.

Une différence majeure, cependant, existe ou devrait exister entre la langue schtroumpf et la langue du droit : l'objectif fondamental d'intelligibilité de la loi devrait, au moins lorsque la loi se veut écrite directement pour le justiciable, exiger une réflexion approfondie sur la compétence à la fois référentielle, culturelle et intertextuelle attendue du destinataire ; or cette étape, au moins en France, est bien souvent oubliée lorsqu'il s'agit de produire concrètement la loi.

C'est une autre différence essentielle entre la langue schtroumpf et la langue du droit qui rend cette ambition intrinsèquement complexe : la langue du droit est une langue technique[10]. Or les langues techniques se construisent à l'écart des langues naturelles, notamment pour éviter leur polysémie ; les langues techniques sont ou devraient être monosémiques. De plus, elles sont conçues pour simplifier et fluidifier le travail des professionnels qui ont à les connaître. Dans ce contexte, leurs régularités, leurs réflexes rédactionnels doivent faciliter la lecture. « Schtroumpfer » un mot, c'est le substituer dans un contexte où le mot qu'il remplace est évident, comme le disait Eco. Mais évident pour qui ? Comme l'écrit le linguiste,

> Supposons qu'un locuteur français de culture moyenne entende un poète schtroumpf déclamer *Je suis le schtroumpf, - le schtroumpf, - l'inschtroumpfé* ou un locuteur italien de culture moyenne entende un poète schtroumpf italien, cette fois-ci, réciter *Nello schtroumpf dello schtroumpf di nostra schtroumpf*. Certainement saisiraient-ils celui-là le renvoi à Nerval et celui-ci à Dante. En

[10] Ou « langue de spécialité » (GÉMAR, 1990).

> entendant *To schtroumpf or not to schtroumpf*, ils saisiraient sans doute le renvoi à Shakespeare. Mais ils seraient peut-être perplexes en entendant *Schtroumpf is the schtroumpfest schtroumpf*, parce qu'ils n'auraient jamais pré-entendu l'*April is the cruellest month* de T. S. Eliot. (ECO ,1997, 4.7.2)

2. Le Schtroumpf civil sert (peut-être) à quelque chose

Si l'on transpose cette idée au texte juridique, la clarté qu'il recherche, son intelligibilité appuyée entre autres sur ses formules récurrentes ne relèvent-elles pas d'une fausse évidence qui présuppose une culture (juridique) partagée entre l'auteur et le lecteur[11] ?

La question n'est pas entièrement tranchée ; si l'écart entre langue ordinaire et langue du droit est aisément perceptible et mesurable par le vocabulaire, l'écart entre la syntaxe de la langue juridique et celle de la langue générale n'est jamais véritablement ressaisi. Or comme le montre l'exemple d'Eco, la syntaxe est déterminante pour le lecteur dans sa tâche de décryptage. D'où l'idée, au moyen de la *schtroumpfisation*, de manifester ces régularités dans le texte juridique : en employant la langue schtroumpf dans sa version systématique proposée par Eco, dans quelle mesure le texte juridique est-il *schtroumpfable*, c'est-à-dire habité par des motifs syntactiques de la langue générale si reconnaissables qu'ils sont restituables par le profane ?

2.1. Stratégies de traduction automatique vers le schtroumpf : approche par la loi de Zipf

Deux approches peuvent être convoquées pour tenter de quantifier ce phénomène : la première repose sur une approche statistique classique. L'usage de proverbes dans *Les Schtroumpfs* le montre : au-delà d'une certaine fréquence, une expression est si identifiable qu'elle reste restituable même avec très peu de vocabulaire subsistant (*schtroumpf qui schtroumpfe n'amasse pas schtroumpf*). G. K. Zipf avait montré que dans un texte en langue naturelle, la fréquence des termes employés décrit une loi de puissance : les termes les plus utilisés sont exponentiellement plus fréquents que ceux qui le sont moins (ZIPF, 1929, 1932). La « courbe de Zipf » donne une représentation graphique de ce phénomène, ici par exemple pour le Code civil québécois :

[11] Ou pour exprimer le problème comme Eco dans *Kant et l'ornithorynque* : « La langue schtroumpf répond donc aux règles d'une linguistique du texte, où le sens dépend de l'identification du *topic* textuel. Il est vrai que tout texte est une machine paresseuse qui réclame une active coopération interprétative de la part de son destinataire, ce qui semblerait nous inviter à faire des textes en schtroumpf : mais notre collaboration est possible parce que nous nous reportons à l'univers de l'intertextualité ; de même, nous pouvons comprendre le schtroumpf parce que chaque locuteur utilise le terme schtroumpf et ses dérivés seulement et toujours dans des contextes où une phrase de ce genre a déjà été prononcée. »

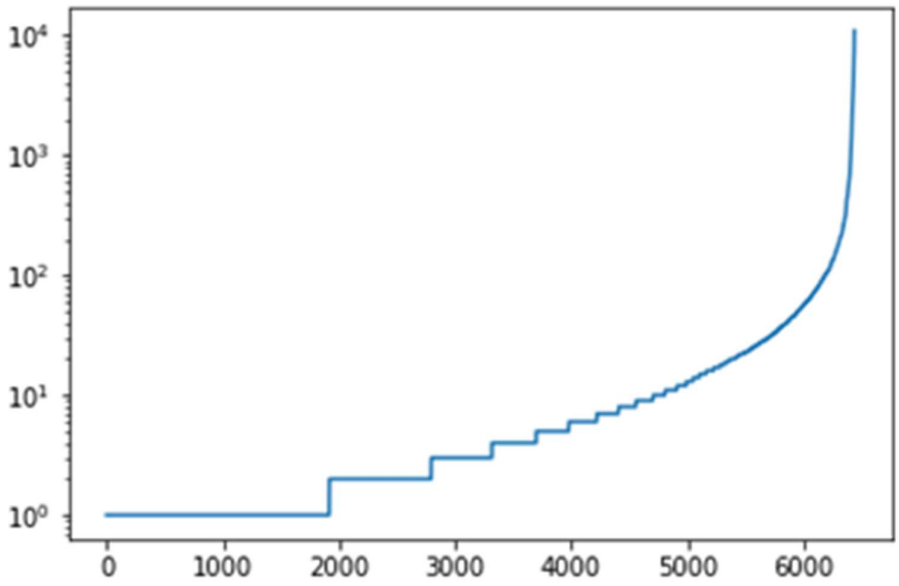

Figure 5 - Rang vs. occurrences des 1-grammes, Code civil du Québec

La même règle empirique se vérifie-t-elle pour des séquences de mots ? L'expérience, menée sur les séquences de quatre mots du Code civil québécois, semble le prouver :

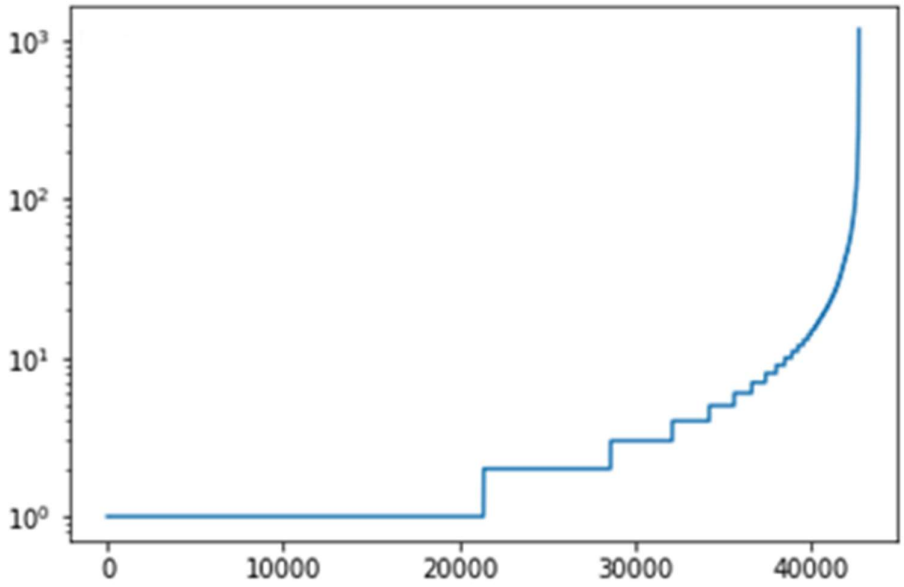

Figure 6 - Rang vs occurrences des 4-grammes, Code civil du Québec

Il suffirait alors de rapprocher les fréquences des séquences de quatre mots (ou 4-grammes) dans un corpus représentatif de la langue générale (dont

la constitution est moins aisée qu'il paraît) ou représentatif d'un niveau de maîtrise de la langue, des 4-grammes du Code civil. Si l'on trouve empiriquement à quel point de la courbe de Zipf la syntaxe permet de pallier la *schtroumpfisation* du vocabulaire, toutes les séquences plus fréquentes sont réputées restituables, donc schtroumpfables.

2.2. Approche fondée sur l'apprentissage machine

Une autre approche consiste à mettre en œuvre des techniques dites d'apprentissage machine. Un modèle de langue est construit en observant les régularités d'une très grande quantité de texte ; le modèle peut ensuite essayer de prédire les parties manquantes d'une phrase en s'appuyant sur les parties conservées. Comme le lecteur, ces modèles sont des machines à contexte ; mais contrairement à lui, ils sont probablement incapables de passer d'une série de contextes à un sens général[12]. Leur succès ne passe pas par une compréhension conceptuelle des textes qu'ils traitent, mais par une extraordinaire efficacité dans l'art de rapprocher les contextes, d'examiner les cooccurrences à une échelle industrielle tout incapables qu'ils sont de raisonnement logique. La démarche revient, pour reprendre l'expression d'Eco, à brancher une véritable « machine paresseuse » interprétative à la machine paresseuse qu'est pour lui le texte, raison suffisante pour privilégier cette approche.

Pour mener à bien l'expérience, nous avons choisi le modèle ouvert « French Europeana ELECTRA » (CLARK, LUONG, LE QUOC *et al.*, 2019), développé par l'équipe numérique de la bibliothèque d'État bavaroise. Appuyé sur les textes en français du XVIIIe au XXe siècles accessibles à la bibliothèque numérique Europeana, il a été entraîné sur 11 milliards de mots (à titre de comparaison, le Code civil du Québec en compte environ 184 000) ; les Codes civils français et québécois ne font pas partie du corpus d'entraînement. On peut faire l'hypothèse que les régularités sémantiques et syntaxiques qu'il a intégrées sont comparables à celles d'un humain à la solide culture générale francophone.

On demande ensuite au modèle de langue de prédire chaque mot des Codes civils appartenant aux catégories *nom*, *verbe* ou *adverbe* ; si le mot considéré le plus probable par le modèle est le bon, il est considéré aisément restituable et est donc *schtroumpfé*.

2.3. *Stroumpfisation* et test de closure : un nouveau type de mesure de la lisibilité, voire d'une forme d'intelligibilité ?

Cette technique, qui consiste à effacer un mot d'un texte et à essayer de le deviner, ressemble fortement dans sa démarche au test de closure,

[12] Même si les dernières familles publiées de modèles de ce type, notamment la famille des GPT, peut en donner l'illusion.

développé par Taylor en 1953 ; l'idée était de mesurer la lisibilité en demandant à des élèves de compléter un texte dont un mot sur cinq avait été effacé. Diverses variantes ont ensuite été proposées, avec pour ambition de mesurer la compréhension du texte et/ou sa lisibilité (GAGATSIS, 1985, FORTIER, 1982). La littérature pédagogique a abondamment discuté ce que mesure exactement le test de closure ; si un certain consensus semble se dégager pour considérer qu'il est efficace pour prédire la lisibilité d'un texte, le débat semble encore ouvert sur la question de savoir s'il peut en prédire l'intelligibilité. Une question fréquemment posée est de savoir si c'est bien l'intelligibilité du texte ou l'intelligence du lecteur qui est mesurée. Le Schtroumpf civil pourrait-il servir à quelque chose ? Avec cette technique (utiliser un modèle de langue pour prédire le mot effacé), la mobilisation d'une authentique *machine paresseuse* neutralise le critère de l'intelligence, resserrant la donnée produite autour de la prédictibilité contextuelle du mot. Si un texte hypothétique était intégralement prévisible, il serait entièrement schtroumpfé. Si cette prédictibilité (schtroumpfabilité ?) n'épuise pas, loin de là, la question de l'intelligibilité, notoirement difficile à quantifier, elle ouvre des perspectives intéressantes : comment évoluerait-elle si la prédiction était réalisée par d'autres modèles, reflétant par exemple le niveau de compétence linguistique et culturelle d'un élève sortant de l'enseignement secondaire, voire primaire ? Taylor, à propos du test de closure, écrivait qu'il mesurait « le degré de correspondance totale entre les habitudes d'encodage d'émetteurs et les habitudes de décodage des récepteurs ». Le projet burlesque du Schtroumpf civil a au moins cette vertu : rendre visible au législateur, par ses substitutions, la plus ou moins grande familiarité des tournures qu'il emploie pour ses lecteurs ; vérifier, en d'autres termes, qu'ils habitent bien le même monde (pour ne pas dire le même village). On retrouve le paradoxe originel de la langue schtroumpf : plus il y a de *schtroumpfs*, plus la schtroumpf est schtroumpf.

1. Bibliographie

CLARK, K., LUONG, M.-T., LE QUOC, V. *et al.* (2019). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators, [consulté le 8 mai 2022]. https://openreview.net/forum?id=r1xMH1BtvB

DAYEZ, H. (2003). *Peyo l'enchanteur*, Niffle.

DOMINGUEZ LEIVA, A. et HUBIER, S. (2016). *Schtroumpfologies*, Le Murmure.

ECO, U. (1979). Schtroumpf und Drang, *Alfabeta*.

ECO, U. (1979). *Lector in fabula: la cooperazione interpretativa nei testi narrativi*, (s. l.). Bompiani.

ECO, U. (1999). *Kant et l'ornithorynque*, Grasset.

FLÜCKIGER, A. (2007). Le principe de clarté de la loi ou l'ambiguïté d'un idéal. *Cahiers du Conseil constitutionnel*, (21).

FORTIER, G. (1982). Le test de closure : un outil pédagogique. *Québec français*, Les Publications Québec français, 46, p. 78-82.

GAGATSIS, A. (1985). Questions soulevées par le test de closure. *Revue française de pédagogie*, (1), p. 41-50.

GÉMAR, J.-C. (1990). Les fondements du langage du droit comme langue de spécialité. Du sens et de la forme du texte juridique. *Revue générale de droit*, *21*(4), Éditions Wilson & Lafleur, inc., p. 717-738.

GÉMAR, J.-C. (2019). Analyse jurilinguistique des concepts de ''lisibilité'' et d''intelligibilité'' de la loi. *Revue générale de droit*, *48*(2), p. 299-336.

GUY-ECABERT, C. et FLÜCKIGER, A. (2015). La bonne loi ou le paradis perdu? [en ligne]. *LeGes*, (1), [consulté le 29 décembre 2022].

https://leges.weblaw.ch/legesissues/2015/1/2015121-45.html

MALAURI, E. P. (2005). L'intelligibilité des lois. *Pouvoirs*, *3*(114), Le Seuil, p. 131-137.

SARGOS, P. (2008). La causalité en matière de responsabilité ou le "droit Schtroumpf". *Recueil Dalloz Sirey*, *28*, p. 1935-1941.

TAYLOR, W. L. (1953). Cloze procedure: a new tool for mesuring readability. *Journalism Quarterly*.

ZIPF, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in classical Philology*, *40*, Harvard University Press.

ZIPF, G. K. (1932). Selected studies of the Principle of Relative Frequency. *Language*, Cambridge, Mass.