

SERG – Grupo de Pesquisa em Engenharia Semiótica do
Departamento de Informática da PUC-Rio

Série de Notas do EMAPS
Ética e Mediação Algorítmica de Processos Sociais

Leitura Direta e Leitura Informatizada com o Voyant

EMAPS NOTAS #05

Clarisse Sieckenius de Souza
www.inf.puc-rio.br/~clarisse

Abril de 2024

Como citar este documento: de Souza, C. S. (2024) **Leitura Direta e Leitura Informatizada com o Voyant**. *EMAPS-Notas #05* 10 p. Rio de Janeiro, RJ - Brasil: SERG, Departamento de Informática, PUC-Rio, 2024. 10p.

URL da versão impressa: <http://www.hcc.inf.puc-rio.br/EMAPS/userfiles/downloads/Notas-deSouzaEstudoVoyant2024.pdf>

URL da versão interativa: http://www.hcc.inf.puc-rio.br/EMAPS/Voyant?Um_Estudo_Completo

CLARISSE DE SOUZA © 2024 BY-NC-ND 4.0

Leitura Direta e Leitura Informatizada com o Voyant

Clarisse Sieckenius de Souza

Abril de 2024

Resumo:

Esta Nota do EMAPS ilustra e levanta questões sobre um processo de exploração da suíte ferramentas para análise textual informatizada Voyant-Tools. Todas as etapas do processo podem ser experimentadas e exploradas pelos próprios leitores e leitoras, graças a uma característica poderosa e distintiva do Voyant (como nos referimos à suíte neste texto): sua disponibilidade para livre acesso no endereço <https://voyant-tools.org/>. A pergunta-guia do estudo, intencionalmente narrado em primeira pessoa, como sujeito da leitura, é: Como a utilização desta tecnologia transforma a experiência de leitura e análise de textos? Como referência básica, tomamos o livro dos próprios idealizadores do Voyant (Rockwell e Sinclair (2016)) e os trabalhos de Franco Moretti (2013, 2017) sobre Leitura a Distância. As questões levantadas neste estudo tornam-se objetos de investigação que exploraremos em estudos futuros.

Palavras-Chave:

Voyant-Tools, Humanidades Digitais, Retórica Digital, Hermenêutica Digital

Referências

Moretti, F. (2013). *Distant Reading*. New York: Verso.

Moretti, F. (2017). Patterns and interpretation. Stanford LitLab Pamphlet.

Rockwell, G. e Sinclair, S. (2016). *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: The MIT Press.

Leitura Direta e Leitura Informatizada com o Voyant

Clarisse Sieckenius de Souza, Departamento de Informática, PUC-Rio

Referência para Citação

Clarisse Sieckenius de Souza, *Leitura Direta e Leitura Informatizada com o Voyant*. EMAPS-Notas #5 10 p. Rio de Janeiro, RJ - Brasil: SERG, Departamento de Informática, PUC-Rio, 2024. 10p. [Versão interativa aqui](#).
URL da versão impressa: <http://www.hcc.inf.puc-rio.br/EMAPS/userfiles/downloads/Notas-deSouzaEstudoVoyant2024.pdf>

Resumo

Esta Nota do EMAPS ilustra e levanta questões sobre um processo de exploração da suíte ferramentas para análise textual informatizada [Voyant-Tools](#). Todas as etapas do processo podem ser experimentadas e exploradas pelos próprios leitores e leitoras, graças a uma característica poderosa e distintiva do Voyant (como nos referimos à suíte neste texto): sua disponibilidade para livre acesso no endereço <https://voyant-tools.org/>. A pergunta-guia do estudo, intencionalmente narrado em primeira pessoa, como sujeito da leitura, é: Como a utilização desta tecnologia transforma a experiência de leitura e análise de textos? Como referência básica, tomamos o livro dos próprios idealizadores do Voyant (Rockwell e Sinclair, 2016) e os trabalhos de Franco Moretti (2013 e 2017) sobre *Leitura a Distância*. As questões levantadas neste estudo tornam-se objetos de investigação que exploraremos em estudos futuros.

1. Introdução

Desde o lançamento do [ChatGPT](#) pela OpenAI, em novembro de 2022 (OpenAI, 2023), experimentamos, entre muitas outras coisas importantes, uma revolução nos modos de acesso e produção de conhecimento, tema de meu interesse, onde se enquadra este estudo. Pergunto-me, como Romele (Romele *et al.*, 2020; Romele, 2021), o que seria uma *hermenêutica digital*, para então refletir a respeito de seus impactos sobre as nossas condições de conhecimento, daqui para frente. Tudo isto é "trabalho futuro," em jargão acadêmico. No momento, me preparo para ele dando um primeiro passo, com uma tecnologia que antecede o ChatGPT e que, explícita e deliberadamente, coloca como *sujeito intérprete* de qualquer análise textual, informatizada ou não, o ser humano. Esta premissa é valiosa, no sentido de que os atuais *transformers* e *Large Language Models* que permitem colocar a máquina (ChatGPT) como inteligência artificial capaz de interpretar textos, ela mesma, têm uma ligação tecnológica com ferramentas produzidas por décadas de pesquisa em linguística computacional para análise e geração de textos. Poderíamos nós clarear um pouco da "caixa preta" dos LLMs se entendêssemos melhor como estas tecnologias anteriores ao ChatGPT funcionam? É preciso estudar para responder, e este é meu primeiro estudo.

Este trabalho é intencionalmente narrado em primeira pessoa, uma vez que o seu objeto central é um processo de interpretação de texto por parte de seu agente natural primário: uma pessoa que sabe ler. Faço uma comparação entre, por um lado, uma leitura direta, resultante de meu contato direto com o texto em formato impresso, e, por outro, uma leitura informatizada, resultante de meu contato *mediado* por ferramentas de análise computacional de texto. Para esta finalidade, fiz três escolhas *a priori* no início do estudo. A primeira é sobre o TEXTO a ser lido; a segunda é sobre a ORDEM em que faço a leitura direta e a informatizada; e a terceira é a TECNOLOGIA para a leitura informatizada.

O Texto

O texto escolhido para o estudo foi a mensagem do Papa Francisco (2024) para o Dia Mundial da Paz de 2024. Ela pode ser encontrada online neste endereço: <https://www.vatican.va/content/francesco/pt/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html>. Os principais motivos da escolha são: (1) a ampla divulgação do texto; (2) o texto ser relativamente curto e permitir, a qualquer pessoa letrada no idioma em que está escrito, uma leitura atenta; (3) o texto estar oficialmente traduzido pelo Vaticano em mais de dez línguas; e finalmente (4) a temática do texto, tão próxima de meu interesse maior, apresentado no início desta introdução. Ao apresentar os resultados, faço algumas considerações sobre a adequação deste texto para este estudo.

A Ordem de Leitura

A ordem que escolhi para a leitura foi começar pela leitura direta e depois fazer a leitura informatizada. Ao final das duas, a comparação rendeu constatações e gerou perguntas muito importantes. A primeira delas, obviamente, é: O que aconteceria se a ordem fosse a inversa? Em relação a este texto, já me é impossível responder nas mesmas condições do estudo. Mas, dentre os já mencionados "trabalhos futuros," a busca por esta resposta tem alta prioridade.

A Tecnologia

A tecnologia escolhida foi o Voyant-Tools, uma suíte de ferramentas de análise textual disponibilizada para uso público em <https://voyant-tools.org/>. Graças a ela, entre muitas outras coisas, a publicação deste estudo é completamente pública e interativa, permitindo que todas as etapas do processo que realizei sejam experimentadas e exploradas pelos próprios leitores e leitoras com as ferramentas e recursos indicados nas seções seguintes. Trata-se, por um lado, de uma escolha de conveniência e, por outro, de uma escolha técnica, já que o Voyant oferece inúmeras formas de análise e visualização de textos, sem mencionar o fato de que permite a usuários avançados criar *scripts*, com uma API específica para personalizar e estender as interfaces e interações com o Voyant (a *Spyral*), ou até escrever ou injetar programas em Javascript para adicionar funcionalidades à suíte.

Antes de concluir esta introdução, é importante ressaltar que utilizei como fonte referências bibliográficas específicas, que influenciaram radicalmente o que fiz. A primeira delas é o livro *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, onde os idealizadores do Voyant, Geoffrey Rockwell e Stéphan Sinclair (2016), apresentam sua própria visão sobre o que a tecnologia é, seu valor, suas propriedades, funções e limitações. A segunda é um conjunto de trabalhos de Franco Moretti (2007, 2013, 2017), sobre seu conceito de *"distant reading"* e as oportunidades que ele traz para os estudos no campo das Humanidades Digitais.

Nas próximas seções apresento o método que usei neste estudo, os resultados produzidos e a minha avaliação sobre eles. Em uma seção final alinhavo conclusões e falo de trabalhos futuros.

2. O Estudo: Método, Processo, Resultados e Avaliação

Nesta seção apresento etapas, resultados e minha avaliação do estudo, incluindo acesso a ferramentas do Voyant. Além das ferramentas, também apresento *links* para arquivos importantes, caso o(a) leitor(a) deseje explorar caminhos alternativos aos que explorei. Porém, antes de passar aos detalhes de execução do estudo, convém ressaltar o seu **objetivo específico**. Dentro do quadro maior da busca pela resposta para se poderíamos clarear um pouco da "caixa preta" dos LLMs se entendêssemos melhor como tecnologias anteriores ao ChatGPT funcionam, este estudo tem por objetivo **clarear meu entendimento sobre o Voyant**.

2.1. Método

O método elaborado para realizar o estudo privilegia dois pontos centrais: a escolha de uma estratégia *comparativa* do que é uma leitura direta e uma leitura informatizada, com auxílio do Voyant; e a *seleção de um objeto de comparação*, que represente "a minha leitura" em cada caso. Começando por este último, escolhi o **RESUMO DO TEXTO** como base da comparação. Ao final de cada leitura, que resumo da mensagem papal, feito por mim, representa satisfatoriamente o meu entendimento daquele texto?

A estratégia de comparação, porém, é bem mais complexa do que pareceu a princípio. Não se trata de fazer duas leituras, produzir dois resumos e compará-los ao final. Por um lado, já tendo um entendimento prévio do texto, resultante da primeira leitura direta, seria impossível esquecer tudo o que já sabia e escrever um segundo resumo, resultante da leitura informatizada, sem qualquer ingrediente de entendimento formado a partir da primeira leitura. Neste sentido, o "objeto de comparação" permanece, mas a comparação ganha matizes distintos. Por outro, ao invés de a comparação se processar por um processo de cotejamento e análise de diferenças, fiz dela um processo de **ANÁLISE E REVISÃO**. Depois de concluir a leitura informatizada com o Voyant, retornei ao resumo feito ao final da leitura direta e procurei analisar: (a) que partes daquele texto "não apareceram" na minha leitura informatizada (e por quê); e (b) que tipo de ajustes eu poderia fazer ao resumo original para gerar uma segunda versão melhorada dele (e qual seria ela).

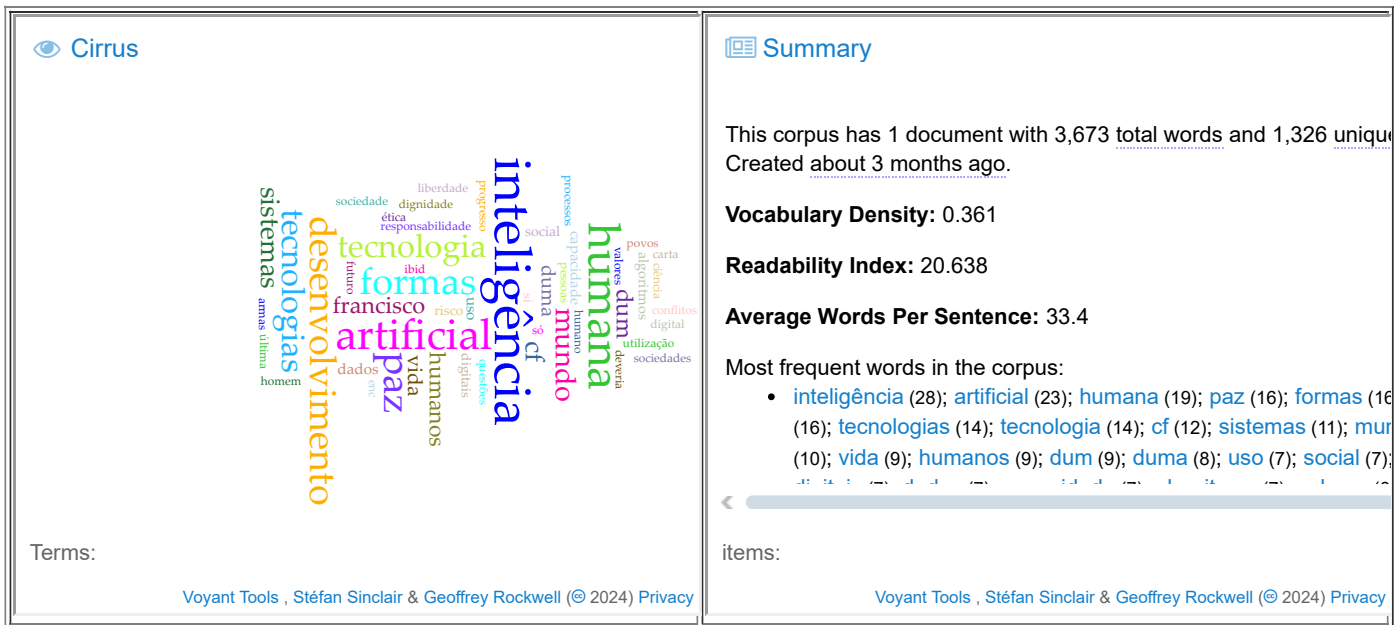
Os passos do método foram, portanto:

1. Fazer uma leitura direta da Mensagem do Papa
2. Escrever um resumo que representasse, adequadamente, meu entendimento da mensagem.
3. Enviar o texto com a Mensagem do Papa para o Voyant.
4. Submeter o texto a análises com as ferramentas do painel básico do Voyant: **Cirrus** (nuvem de palavras); **Trends** (análise de distribuição e frequência de termos ao longo do texto); **Summary** (resumo de metadados quantitativos a respeito de texto); e **Contexts** (lista de concordância para o texto).
5. Explorar estas análises ajustando parâmetros locais ou globais de cada ferramenta.
6. Registrar os resultados obtidos.
7. Analisar e revisar o resumo escrito ao final da leitura direta.

2.2. Pontos Cardeais do Processo

Após a leitura direta da mensagem do Papa Francisco (disponível [aqui](#)), produzi o seguinte resumo:

A mensagem do Papa Francisco para o Dia Mundial da Paz de 2024 é uma reflexão das relações da Inteligência Artificial (IA) com a paz, a essência e a dignidade humana. Diante das promessas e riscos que este tipo de tecnologia digital oferece para a humanidade, o Papa nos exorta a promover e manter, com efetivo engajamento, um olhar crítico e vigilante sobre o desenvolvimento da IA, como indivíduos e sociedade. Exorta também os governantes e líderes da comunidade internacional a elaborar uma regulação normativa e vinculante quanto à produção e uso da IA, nos mais variados campos de aplicação. Para justificar o destaque dado à IA e lastrear o chamado planetário que faz, não apenas aos católicos, mas a todos os homens e mulheres de boa vontade do planeta, o Papa Francisco trata especificamente de oito pontos críticos. Primeiramente, fala que o progresso tecnológico traz oportunidades, mas também traz riscos, sobre os quais é necessário refletir. Em um segundo ponto, o Papa faz uma densa reflexão sobre o fato de que as tecnologias não são neutras, mas trazem embutidos nelas os valores, interesses e objetivos daqueles que as produzem. Aludindo ao frequente uso plural do termo ("inteligências artificiais"), o Sumo Pontífice demarca uma diferença fundamental entre nós, humanos, e as máquinas inteligentes. Não somos ou temos inteligências fragmentárias, que usamos apenas para isto ou aquilo. O terceiro, quarto, quinto e sexto pontos abordam riscos e questões da IA já muito discutidas hoje em dia. Estão incluídos na discussão itens como: o desafio filosófico que máquinas que aprendem oferecem à nossa capacidade de atingir a verdade; o papel dos limites, não apenas numa perspectiva humanista (cristã), onde a vida é finita e a plenitude é uma dádiva só apreciada contra o pano de fundo da finitude, mas também numa perspectiva relacionada à ânsia de superar limites técnicos, econômicos e políticos, levando ao risco da concentração de poder e riqueza na mão de poucos, além da perda de postos de trabalho e suas consequências; os desafios éticos e sociais relativos à delegação, para IAs, de tomadas de decisão e à permissão para que influenciem o comportamento humano, sem qualquer mediação e vigilância; o desenvolvimento e uso de armas inteligentes em uma guerra cada vez mais distante da realidade trágica das populações atingidas; e finalmente a necessidade de se orientar a pesquisa e a técnica, de forma interdisciplinar focada na ética de algoritmos, para o desenvolvimento de aplicações de IA que aliviem as desigualdades e o sofrimento humano em escala global. O sétimo ponto ressalta a urgência de se educar os jovens e preparar os adultos para manter um olhar crítico e vigilante sobre a tecnologia que usam, que desejam, e que se lhes oferece, com destaque para o papel das escolas, universidades e poder público nesta questão. Finalmente, o oitavo ponto convoca a comunidade internacional a trabalhar unida, engajada e desde já, pela regulamentação jurídica do desenvolvimento técnico e uso da IA. A mensagem se encerra afirmando que é responsabilidade de todos, e não apenas de uns poucos, refletir e agir diante do quadro de desenvolvimento da tecnologia digital de hoje. O Papa conclui com votos de que a IA não venha agravar os já demasiados conflitos e injustiças sociais, mas que possa ser usada para promover a dignidade humana e a paz.



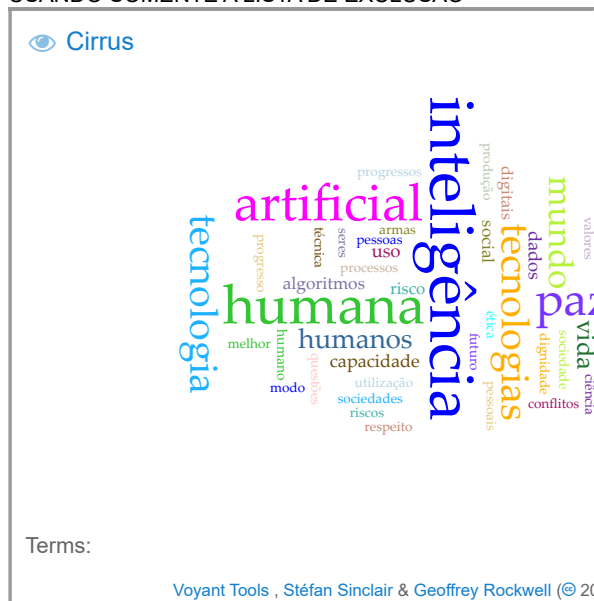
Na ferramenta **Cirrus**, o controlador de número de termos (**slider Terms**, no canto inferior esquerdo) está na posição *default*. Se você arrastar o controle para a direita, vai aumentar o número de termos que aparece na nuvem de palavras, por exemplo. Veja o efeito! Já na ferramenta **Summary**, alteramos o valor *default* no controlador de palavras cuja frequência é apresentada no campo "Most frequent words in the corpus" (**slider items**, no canto inferior esquerdo) para 59. Meu objetivo é mostrar como estão sendo analisadas palavras que não têm o menor interesse para nossa leitura do texto. É, por exemplo, o caso de *cf* (12), *dum* (9) e *só* (6), entre outras.

Atenção: Como o processamento do resultado de algumas análises com parâmetros diferentes do *default* pode demorar (porque é computacionalmente pesado) ou ser indevidamente interrompido no servidor público do Voyant-Tools (porque o tráfego está intenso, por exemplo), caso você não esteja vendo a visualização esperada ou aqui descrita, mexa nos controles da visualização (o que deve surtir efeitos em alguns segundos), ou então exporte a visualização em questão para outra aba ou janela de seu navegador. Uma das duas soluções costuma funcionar corretamente. Caso não funcione, provavelmente há um problema do servidor do Voyant. Todas as análises aqui apresentadas, até a data da publicação final desta nota, foram devidamente testadas e conferidas. Lamento se, com o tempo, ou com mudanças de plataformas, alguma parte deste trabalho ficar inacessível.

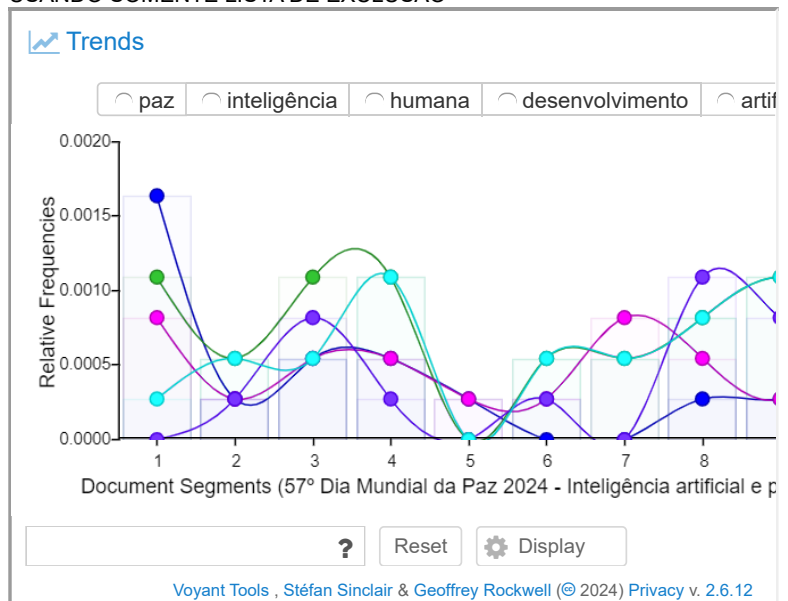
Para refinar a análise, excluindo palavras que não carregam teor semântico relevante, [criei uma lista de exclusão](#) a partir da lista de **stopwords** que o Voyant oferece para a língua portuguesa. Fiz isto manipulando dados exportados para uma planilha EXCEL ([saiba mais sobre exportações de dados no Voyant](#)). A lista final de exclusão pode ser examinada [neste endereço](#). Similarmente, examinando os termos da mensagem, elaborei uma lista de inclusão, que pode ser vista como uma fonte de indexação do texto, uma espécie de *lista estendida de palavras-chave*. Ela pode ser encontrada [neste endereço](#).

O efeito do uso destas listas em algumas ferramentas de visualização do Voyant é bem interessante. Veja alguns exemplos abaixo.

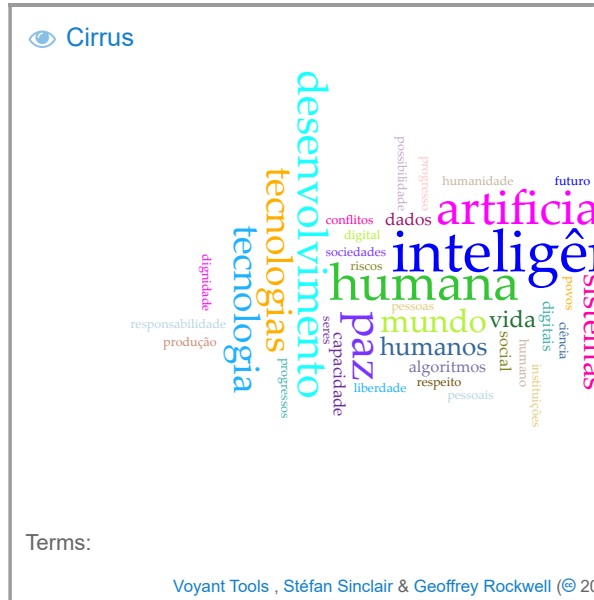
USANDO SOMENTE A LISTA DE EXCLUSÃO



USANDO SOMENTE LISTA DE EXCLUSÃO



USANDO SOMENTE A LISTA DE INCLUSÃO



É bom notar que, entre as ferramentas básicas para análise de corpus e documentos no Voyant, a única onde faz sentido utilizar **somente** a lista de inclusão (i.e. apenas palavras "indexadoras" do conteúdo central do documento) é a **Cirrus** (nuvem de palavras). Nas outras, não apenas a lista de inclusão não é um parâmetro de análise, como também o cancelamento do uso da lista de exclusão faz a ferramenta voltar a tratar palavras desprovidas de teor semântico relevante como se fossem relevantes. Veja em: <https://voyant-tools.org/?corpus=c26037787d03fae83ffc44803844456&stopList=&whiteList=keywords-85d86666e584c0d5408b3bfdced6f3cf&view=Trends> o que acaba acontecendo.

O ponto mais produtivo no meu processo de análise foi examinar, conjuntamente, os resultados das ferramentas **Collocates** e **Contexts**, tendo em conta uma lista estendida de inclusão, que podemos encontrar [aqui](#). Esta última lista contém termos que carregam algum teor semântico relevante, embora nem todos se prestem a ser indexadores (a lista de termos candidatos a indexadores é [esta](#)).

Com a **Collocates**, é possível examinar que termos ocorrem mais ou menos proximo uns dos outros. A variação de distância entre os termos pode ser alterada com o movimento do **slider context**, para a esquerda (limite 1) ou direita (limite 30). Este exercício, associado à indicação de quantas vezes o padrão é encontrado no texto, contribui para um mapeamento de potenciais "temas" da mensagem.

Voyant Tools	
Collocates	
Term	Collocate
inteligência	artificial
inteligência	formas
inteligência	utilizada
inteligência	sistemas
inteligência	humana
inteligência	deveria
inteligência	visar
inteligência	utilizam
inteligência	utilização
inteligência	uso
<input type="text" value="inteligência"/>	<input type="text" value="356 context"/>

Voyant Tools , Stéfán Sinclair & Geoffrey Rockwell © 2024 Privacy v. 2.6.12

Os "temas candidatos" podem ser procurados com a ferramenta **Contexts**, que permitirá uma inspeção detalhada do contexto de ocorrência. No processo de inspeção, podem ser utilizadas expressões regulares para buscas de termos. O Voyant oferece um conjunto de operadores que representam padrões de sequências alfanuméricas (veja mais [aqui](#)). Abaixo, com a ferramenta **Contexts**, mostro a ocorrência do padrão `progress*|melhor*|"o bem"`

Contexts				
Document	Left	Term	Right	
57º Dia ...	a fazer com que os	pr...	no desenvolvimento de formas de	
57º Dia ...	entre promessas e riscos Os	pr...	da informática e o desenvolvimen	
57º Dia ...	comum. [3] Deste modo os	pr...	notáveis das novas tecnologias di	
57º Dia ...	sofrimentos. Ao mesmo tempo, os	pr...	técnico-científicos, que permitem	
57º Dia ...	aspetos sociais e éticos do	pr...	e da utilização da tecnologia	
57º Dia ...	as leis que promovem o	pr...	humano integral. A inteligência art	
57º Dia ...	poderão ser considerados um verdadeiro	pr...	. [8] A inteligência artificial tornar	
57º Dia ...	sua utilização, para que o	pr...	digital possa verificar-se no	
57º Dia ...	os povos. Assim o próprio	pr...	da ciência e da técnica	
57º Dia ...	votos de paz. 1. O	pr...	da ciência e da tecnologia	
57º Dia ...	Discurso aos participantes no Seminário «	o ...	comum na era digital» (27	

? 19 context expand

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2024) Privacy v. 2.6.12

A partir das explorações de termos do texto, fiz uma **categorização** de padrões lexicais e morfológicos significativos para a mensagem, dados os seus respectivos contextos de ocorrência. As categorias geradas foram as seguintes:

RISCOS E AMEAÇAS

afli* | ameaça* | arma* | bélic* | conflito* | desigual* | econ* | ego* | empreg* | espada* | guerra* | injustiça* | leta* | militar* | necessitad* | perigo* | pobre* | preconceit* | risco* | terrorista* | trabalhador* | trabalho*

PROGRESSO

avan* | crescimento | desenvolvi* | dinâmico* | efici* | expan* | facilit* | fáceis | progre* | "o bem" | melhor*

TECNOLOGIA

"aprendizagem de máquina" | "inteligência artificial" | "inteligências artificiais" | "machine learning" | "nova tecnologia" | "novas tecnologias" | algor* | comput* | dados | digita* | internet | sistemas | tecno*

HUMANIDADE

homem | homens | human* | jove* | mulher* | mundo | pessoa* | povo* | sociedade*

VALORES, ÉTICA, MORAL

"boa vontade" | ambient* | amor* | bem | convivência | dign* | educ* | equitat* | etic* | fratern* | justiça* | justo* | liberdade | moral* | pacif* | pacífic* | paz* | solid* | valor* | étic*

RELIGIÃO

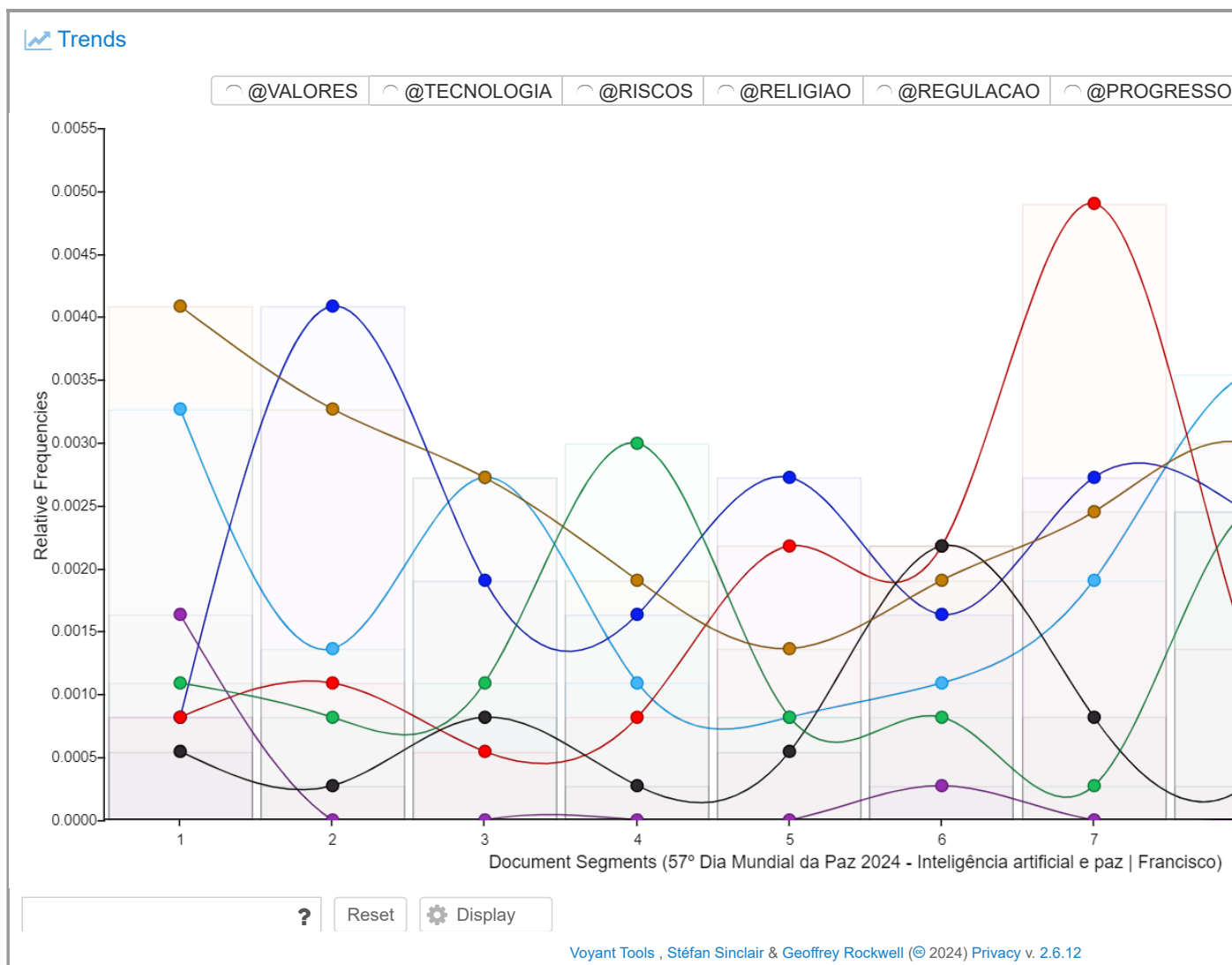
deus | crente* | crist* | divin* | dádiv* | graça | relig*

REGULAÇÃO

reg* | jur* | lei* | lega* | legisla* | respons* | limit* | chefe* | govern* | aten*

A categoria PROGRESSO foi revisada na data de publicação deste documento. Mencionada em outras publicações e cadernos do EMAPS, ela pode conter um conjunto de termos ligeiramente diferente.

Usando a ferramenta **Trends**, abaixo, comparo a distribuição de ocorrência dos termos para as categorias que criei. O uso do símbolo "@" seguido do nome da categoria torna o mecanismo de buscas mais eficiente, para várias ferramentas do Voyant.



Um último destaque dos procedimentos e etapas de análise que devo comentar é como a ferramenta **Trends**, tal como configurada acima, é um ponto de partida para várias interações muito ricas em informação sobre os padrões textuais da mensagem do Papa. Primeiramente, é possível *ligar e desligar* a visualização de cada categoria. Se clicarmos no nome da categoria ao alto do gráfico e ela estiver visível (*ligada*) ela fica invisível (*invisível*), e vice-versa. Um exercício com este recurso pode mostrar quais padrões de distribuição de frequência são mais semelhantes ou mais diferentes. Experimente!

Além deste recurso, é possível controlar a segmentação do texto analisado, isto é, estipular em quantas partes iguais ele vai ser fatiado para fins de visualização. Este controle aparece nas "opções" da ferramenta **Trends**. Um segmento é, portanto, uma fração do documento, fatiada por critérios formais: por exemplo, se um documento é segmentado em 10 partes, cada parte corresponde a 10% do documento. Ou seja, não há uma correspondência com a "estrutura" do documento, a menos que o texto tenha sido formalmente marcado (com uso de XML) para este tipo de segmentação.

No caso da mensagem do Papa, não estruturamos o documento. Portanto, o fracionamento do texto é matemático e o *default* para o número de segmentos é 10. É o que se vê acima. Porém, se começamos a alterar o número de segmentos para visualização, usando o controle **Segments** do diálogo de opções, algumas curiosidades aparecem. Por exemplo, se dividimos o texto em 6 segmentos, algumas curvas formam padrões visuais interessantes, instigando uma leitura próxima para verificar qual o significado do padrão. É, por exemplo, o caso destas curvas:

- @RISCOS junto com @REGULACAO (que convergem)
- @RISCOS junto com @PROGRESSO (que divergem)
- @VALORES junto com @PROGRESSO (que se acompanham)
- @VALORES junto com @RISCOS (que primeiro convergem e depois se defasam)

Para facilitar o exercício, você pode tomar esta URL por base:

```
https://voyant-tools.org/?
corpus=c26037787d03fae83fffc44803844456&categories=9d66388fdf61b0ac00f577146f5c8a4c&view=Trends
&query=@REGULACAO&query=@RELIGIAO&query=@HUMANIDADE&query=@RISCOS&query=@TECNOLOGIA&query=@VALORES&query=@PROGRESSO
&bins=6&mode=document
```

e editar os parâmetros `query=...` para ficarem apenas as categorias que lhe interessam, bem como alterar o valor do parâmetro `bins=...` para explorar novos padrões e combinações.

2.3. Resultados

Feitas as análises da mensagem do Papa Francisco no Voyant, a etapa final do método, para chegar aos resultados finais foi um retorno ao resumo produzido logo após a leitura direta, e analisar: (a) que partes dele "não apareceram" na minha leitura informatizada (e por quê); e (b) que tipo de ajustes eu poderia fazer ao resumo original para gerar uma segunda versão melhorada dele (e qual seria ela). Para responder estas perguntas, meus principais achados com o estudo são listados a seguir.

- Ficou muito evidente, na comparação, como o resumo feito após a leitura direta expressa minha reação ao *pathos* da mensagem do Papa, seu efeito retórico em mim. O vocabulário e as frases no meu texto espelham as escolhas eloquentes do Papa. Em comparação, minha leitura distante foi completamente *semântica*, sem a menor atenção à *retórica* da mensagem.
- Possivelmente como consequência do ponto acima, REFLEXÃO e EXORTAÇÃO foram categorias que escaparam completamente da minha análise com o Voyant, embora sejam o ponto-chave do resumo manuscrito.
- A análise do resumo original também mostrou que incluí nele algumas das minhas "pepitas favoritas" da mensagem. É o caso de a tecnologia não ser neutra, de que a inteligência humana não é fragmentada, de que talvez não possamos distinguir a verdade da falsidade, e assim por diante. Embora eu não acredite que isso seja um problema – ao contrário, pode ser minha contribuição individual para o vasto universo de leituras possíveis que todos os textos têm a oferecer – pode ser problemático em resumos. Meus temas favoritos não foram desenvolvidos na mesma extensão pelo autor. Isso ficou claro na leitura distante.
- A leitura distante também sugere que esta mensagem não era sobre religião. Mesmo que tenha sido escrita pela autoridade suprema da Igreja Católica e referências a documentos religiosos sejam abundantes, a perspectiva deste documento é mais *ética* do que *religiosa*. Isso ficou claro com o uso da categoria @RELIGIAO na ferramenta **Trends**. O *ethos* do texto, o efeito retórico de ter sido escrito por um papa, influenciou ambas as minhas leituras, a próxima e a distante. A influência, no caso desta última, foi ter criado uma categoria @RELIGIAO que, como se pôde ver, tem um padrão de ocorrência muito menos significativo que as demais. O resumo original, em particular, está cheio de referências ao Papa, como o autor da mensagem, apesar de o texto, em si, só fazê-lo em seu cabeçalho e na assinatura no final (excetuada a lista de referências). No entanto, o texto é escrito em primeira pessoa, um traço retórico que impacta a leitura do início ao fim. Uma análise cuidadosa do Voyant (com busca por verbos flexionados em primeira pessoa do singular, por exemplo) poderia pinçar estas ocorrências; resta saber, porém, se a sua frequência apontaria para o seu significado retórico.
- Um último aspecto retórico da mensagem – que pode ter me afetado, mas não percebi – diz respeito às regras e estrutura subjacentes (o *logos*), agradavelmente retratadas nos padrões das curvas produzidas pela ferramenta **Trends**, ilustradas acima. Aqui, o valor de uma abordagem formal e distante foi evidente. A maneira como as categorias temáticas foram distribuídas e reiteradas, em consonância, alternância, ou contraponto, umas em relação às outras, é uma *micro-retórica* que não percebi na leitura direta da mensagem. Acredito que ela possa ter contribuído significativamente para tornar a mensagem mais clara e convincente, mas não investiguei a questão. Talvez a análise de padrões formais traga efetivamente um tipo de entendimento que a leitura direta não necessariamente detecta. Eis, então, uma janela aberta para o *logos* retórico, que o Voyant poderá ajudar a explorar.

Minha resposta corrente para a pergunta (a) ainda precisa dizer por quê certos significados da mensagem não apareceram na análise informatizada. Acredito que seja porque não estão expressos em padrões morfo-lexicais recorrentes. Não encontrei na análise uma expressão quantitativa da presença destes aspectos, o que não quer dizer que ela não exista. No entanto, quer pelo menos dizer que não é simples percebê-la. Já a resposta para a pergunta (b), referente a revisão do resumo motivada pela análise informatizada, começa por apresentar um resumo *efetivamente revisado*, apresentado abaixo (se quiser recuperar o original, [veja aqui](#)). As partes revisadas estão marcadas. Quando porções do resumo original foram cortadas, insiro o símbolo #; quando foram reescritas; mudo a cor da fonte para verde.

A mensagem do Papa Francisco para o Dia Mundial da Paz de 2024 é uma reflexão das relações da Inteligência Artificial (IA) com a paz, a essência e a dignidade humana. Diante das promessas e riscos que este tipo de tecnologia digital oferece para a humanidade, o Papa nos exorta a promover e manter, com efetivo engajamento, um olhar crítico e vigilante sobre o desenvolvimento da IA, como indivíduos e sociedade. Exorta também os governantes e líderes da comunidade internacional a elaborar uma regulação normativa e vinculante quanto à produção e uso da IA, nos mais variados campos de aplicação. Para justificar o destaque dado à IA e lastrear o chamado planetário que faz, não apenas aos católicos, # o Papa Francisco trata especificamente de oito pontos críticos. Primeiramente, fala que o progresso tecnológico traz oportunidades, mas também traz riscos #. Em um segundo ponto, o Papa faz uma densa reflexão sobre o fato de que as tecnologias não são neutras, mas trazem embutidos nelas os valores, interesses e objetivos daqueles que as produzem. # O terceiro, quarto, quinto e sexto pontos abordam riscos e questões da IA já muito discutidas hoje em dia. Estão incluídos na discussão itens como: o desafio filosófico que a IA oferece à nossa capacidade de atingir a verdade; o papel dos limites, não apenas numa perspectiva humanista, # mas também numa perspectiva # econômica e política associada à concentração de poder; os desafios éticos e sociais de se delegar às máquinas tomadas de decisão sobre pessoas e permitir que direcionem o comportamento humano; o desenvolvimento e uso de armas inteligentes em guerras cada vez mais distantes da realidade trágica das populações atingidas; e finalmente a necessidade de se orientar a pesquisa e a técnica, de forma interdisciplinar e focada na ética de algoritmos, para o desenvolvimento de aplicações # que aliviem as desigualdades e o sofrimento humano em escala global. O sétimo ponto ressalta a urgência de se educar os jovens e preparar os adultos para manter um olhar crítico e vigilante sobre a tecnologia que usam, que desejam, e que se lhes oferece, com destaque para o papel das escolas, universidades e poder público nesta questão. Finalmente, o oitavo ponto convoca a comunidade internacional a trabalhar unida, engajada e desde já, pela regulamentação jurídica do desenvolvimento técnico e uso da IA. A mensagem se encerra afirmando que é responsabilidade de todos, e não apenas de uns poucos, refletir e agir diante do quadro de desenvolvimento da tecnologia digital de hoje. O Papa conclui com votos de que a IA não venha agravar os já demasiados conflitos e injustiças sociais, mas que possa ser usada para promover a dignidade humana e a paz.

Como se pode ver, os cortes foram muito mais recorrentes do que as reformulações da escrita. Considerando que o resultado da análise informatizada com o Voyant orientou fortemente minha atenção para o teor *informativo* do texto, ficou clara na sua revisitação a presença de informações que são perfeitamente dispensáveis, tanto do ponto de vista de uma caracterização semântica, quanto retórica. O resultado da revisão é um texto muito mais conciso (cerca de 20% menor).

2.4. Avaliação

Minha avaliação destes resultados gira em torno de alguns pontos centrais.

O objeto lido

O objeto lido diretamente e o objeto lido através de uma ferramenta como o Voyant não é o mesmo. Embora Moretti (2013, 2015) insista nesta questão, a experiência vivida das duas leituras é a forma mais eficaz de perceber a diferença. Ao invés de me perguntar se em um e outro caso temos leitura e interpretação *legítimas*, o que pode direcionar a discussão no sentido de qualificar uma e desqualificar a outra, me interessa mais indagar sobre o fato de que **um mesmo objeto material**, quando lido com "um par de óculos" e quando lido "com o Voyant" se transforma em outra coisa. Como o sujeito das duas leituras e subsequentes interpretações *sou eu*, em ambos os casos, e também, como mostrou o estudo, a construção significados, nas duas situações, é diferente e leva a conclusões diferentes, parece claro que cabe uma investigação de peso sobre o efeito desta nova espécie de *prótese cognitiva* (v. Santaella, 2019), mediadora entre um sujeito intérprete e um objeto material específico (que, no entender de Santaella (2019), pode ser, por sua vez, uma outra prótese cognitiva).

A leitura

Ao falarmos do objeto lido nos referimos a um par de óculos, ao Voyant, e próteses cognitivas. De que tipo de processo participam estes *mediadores*? Invocando novamente o conceito de leitura distante, ou a distância (*distant reading*) de Moretti (2013, 2017), o processo de que estes mediadores participam não é o mesmo. O problema, como discute Drucker (2017), entre outros autores, é o que entendemos por "leitura." A autora chama a atenção para o fato de que, muito equivocadamente, alguns pesquisadores acreditam que a "leitura" informatizada é *objetiva*, em contraste com a leitura direta, que é situada, influenciada por estados mentais e afetos retóricos, crenças, malentendidos, e tanto mais. A experiência com o Voyant mostra claramente que o intérprete final, em uma "leitura" ou outra, é o mesmo. Fui eu, nos dois casos, a intérprete da mensagem do Papa Francisco, lida diretamente de um endereço na Internet, ou pelas lentes das diversas ferramentas do Voyant. Portanto, a questão da objetividade é uma falácia. Mas ela nos dá pistas de que há algo a ser melhor conhecido e definido, a que tortamente podemos ainda estar chamando de "leitura" e qualificando de "objetiva." Há oposições e diferenças claras. E aqui cabe citar o trecho final do artigo de Drucker (2017), onde com muita clareza ela diz que, a seu ver, não estamos falando nem de leitura, nem de qualquer coisa distante:

We can engage in critical conversation about the differences between mechanistic and hermeneutic work as they inform visualizations and data production and influence cultural practices. Distant reading, when properly understood, is neither mechanistic nor hermeneutic. Its literalness makes it the closest form of reading imaginable. What distant reading lacks is distance. That distance is critical; it is the space between the literal text and the virtual text, between the inscriptional, notational surface and the rhetorical, cognitive effect that produces a text. (p. 634)

Ou seja: temos diante de nós uma tarefa importante e interessante.

A interpretação

Uma vez que **um mesmo objeto material** se bifurca em objetos de interpretação distintos, que resultarão de processos de interpretação distintos, a discussão sobre o que é interpretação, como resultado estabilizado de um processo hermenêutico humano (lembrando que o sujeito é o mesmo em ambas as situações aqui comparadas) não pode ser trivializada. Por um lado, claro que não é a mesma coisa. Por outro, a pergunta mais relevante foi colocada desde cedo por Moretti: qual a relação entre as duas interpretações? A resposta elegante do autor é a de que a função das duas interpretações é *desafiarem-se, intelectualmente, uma à outra* (Moretti, 2013). Em outras palavras, uma gera perguntas para a outra e, assim, estimula o avanço do processos hermenêuticos da outra, para o que, no campo da semiótica Peirceana, se chama de semiose contínua, ou ilimitada (The Peirce Edition Project, 1998).

Das incertezas sobre o estudo

São várias as incertezas sobre este estudo, mas antes de falar sobre elas é importante recuperar o contexto em que ele foi realizado. Trata-se de um primeiro passo, cujo objetivo principal é ensinar um entendimento inicial sobre o que é um processo hermenêutico informatizado, no contexto de uma jornada muito mais longa e tortuosa em busca de respostas sobre o que seria uma hermenêutica digital e quais os seus impactos sobre nossas condições de conhecimento, com o advento de tecnologias como o ChatGPT. Como primeiro passo, houve todo o aprendizado que apresentamos acima, mas também fragilidades claras.

A primeira delas é ter usado um texto curto, onde "padrões recorrentes" e "frequências" tornam-se conceitos discutíveis. Mais discutíveis ainda são as inferências a partir deles. No microcosmo da análise, padrões e frequências podem ser quase fortuitos. E mesmo que não sejam, a relevância de seu significado (quando não o próprio significado) pode ser muito questionável. Para atenuar o impacto desta fragilidade sobre o resultado do estudo e as conclusões traçadas na próxima seção, vale lembrar que o objetivo do estudo é *epistêmico*, não *científico*. Não estou fazendo alegações sobre o que o Papa disse ou não disse em sua mensagem; apenas examino aquilo que entendi do que ele disse, quando fiz uma leitura direta, quando fiz uma análise informatizada do texto, e finalmente quando integrei as duas coisas.

Uma outra fragilidade do estudo, ainda neste estágio epistêmico inicial, é o meu conhecimento ainda "intermediário" do Voyant. Por exemplo, assinalo nos resultados que não consegui encontrar uma expressão quantitativa para certas características retóricas presentes na minha leitura direta da mensagem do Papa. É possível que esta expressão exista, que eu a encontre à medida que descobrir mais recursos do Voyant e que entender melhor certas nuances daqueles que já conheço. Esta lacuna técnica com a ferramenta é, sem dúvida, um problema. No entanto, acredito que o caminho de exploração do Voyant seja, como legítimo processo de semiose (The Peirce Edition Project, 1998), um trajeto aberto. Sempre haverá coisas que ainda não descobri no Voyant (até porque ele é um ambiente reprogramável, mas não só por isto). A questão é estabelecer pragmaticamente um critério de suficiência para o conhecimento técnico da ferramenta, quando se quer tirar conclusões mais profundas e de maior alcance. De fato, não sei que critérios poderiam ser invocados. Tenho certeza de que eles existem e temo se não houver suficientes discussões sobre eles, especialmente no que diz respeito à validade de pesquisas científicas instrumentadas pelo Voyant.

Uma terceira fragilidade do estudo, uma vez mais ainda que considerando seus propósitos epistêmicos, é a falta de contraste *com uma outra interpretação humana*, caminhando em paralelo comigo. A falta desta alteridade certamente enviesada e empobrece os resultados do estudo. Na prática, isto corresponde a saber o quanto seria possível aprender se fôssemos ao menos duas pessoas fazendo o mesmo estudo, seguindo o mesmo método, e discutindo entre nós, ao final, o resultado. Uma *microtriangulação* desta ordem poderia gerar *insights* muito mais poderosos sobre todo este processo de sucessivas análises e sínteses, a partir das "leituras," próxima e distante, da mensagem do Papa Francisco.

Uma quarta fragilidade, que é em si suficiente para desqualificar qualquer interpretação *científica* deste estudo é que, justamente por meu maior interesse estar no processo de uso e análise do Voyant, e definitivamente **não estar** na análise das ideias do pontífice sobre inteligência artificial, ou da política do Vaticano em relação ao avanço tecnológico, não dediquei a necessária atenção a depurar cuidadosamente listas de exclusão, à polissemia lexical, aos riscos de interpretação equivocada de visualizações, fora aos eventuais erros de processamento de palavras (por erro de digitação ou ortografia, ilegibilidade de caracteres, etc.). Sem estes e outros cuidados, minhas interpretações a partir de uma "leitura a distância" da mensagem do Papa Francisco com o Voyant é desprovida de qualquer peso. Embora se pudesse supor que, não sendo a análise do texto a finalidade central do estudo, este ponto não é exatamente uma fragilidade, penso que é. É importante, pois o contraste entre o entendimento obtido com duas leituras "rasas" de um texto (pois até a leitura direta também não foi uma análise de discurso digna do nome), com uma ponderação pouco densa sobre seus significados profundos, indica direções e possibilidades acerca das duas espécies de leitura. Mas ainda não leva a nenhuma conclusão de peso.

Haverá, com certeza, várias outras fragilidades, mas as que acabo de listar parecem levantar questões suficientemente complexas, que precisam avançar até que acrescentemos outras a este rol. Apesar delas, no entanto, algumas conclusões são possíveis. Com elas, encerro, na próxima seção, a apresentação deste estudo.

3. Conclusão

O estudo aqui apresentado cumpriu a finalidade de abrir caminho para um entendimento sobre como a mediação tecnológica oferecida pelo Voyant transforma a experiência de interpretação de textos. Apesar de todas as fragilidades apontadas na seção anterior, foi possível perceber claramente as diferenças e contribuições de cada tipo de "leitura," usando ainda os conceitos propostos por Moretti(2013, 2017), para um entendimento integrado de texto. Também foi possível antecipar que, diante de textos mais volumosos, em particular aqueles que se prestam a uma *leitura por consulta*, o Voyant, através de sua ferramenta **Contexts**, pode vir a ser extremamente útil. Não se trata, porém, como já abordado acima, de se fazer uma leitura e não a outra, mas de se trafegar entre uma e outra, trazendo para a leitura direta o benefício da ampla visão de contextos e padrões recorrentes oferecida pelo Voyant e, em contrapartida, trazendo para a análise no Voyant, os elementos de *pathos*, *ethos* e *logos* formados pelo contato direto com o texto do autor. O processo e os resultados de análise ilustraram de forma interessante este diálogo que se torna possível com as iterações das duas leituras.

Há duas características do Voyant que são imediatamente óbvias e dignas de nota, especialmente em face do que os algoritmos de Inteligência Artificial têm feito com textos (escritos por humanos ou por outros algoritmos). A primeira é que o **sujeito hermenêutico do Voyant é humano**. Até mesmo o design de interface do Voyant (Rockwell e Sinclair, 2016) deixa claro que a tarefa de interpretação do que lá está é do usuário. A segunda é que, como discute muito bem Drucker (2017), toda ferramenta tecnológica é produto de um design humano, de uma **interpretação prévia do que é um texto, do que é analisá-lo, do que é interpretá-lo, do que é significado, do que é significativo**, e assim por diante. Os idealizadores do Voyant oferecem inúmeros testemunhos, em múltiplos canais de comunicação (ver por exemplo [estes](#)) além do módulo de ajuda da ferramenta, sobre qual é a sua visão *hermenêutica*. Embora se possa discuti-la, é inegável o valor deste testemunho comparativamente à opacidade de outras ferramentas similares, e mais ainda à opacidade de um ChatGPT.

Uma terceira característica, que não foi explorada neste estudo porque, como mencionado, depende de conhecimentos técnicos avançados, é a *programabilidade* do Voyant, ou seja, a possibilidade de adicionar ou modificar processos de análise automática de texto para que reflitam a forma específica como o sujeito hermenêutico, usuário da ferramenta, quer fazer (presumivelmente porque assim entenda que seja) o seu processo interpretativo. Esta característica, especificamente, parece oferecer um rico caminho de investigação para o que vem a ser, afinal, uma **hermenêutica digital** (Romele *et al.*, 2020 e Romele, 2021). Especialmente quando aliada a uma quarta característica, que discuti na seção 2 -- o fato de que o Voyant nos projeta numa **conversa com quem nos lê**-- a *programabilidade* do Voyant pode permitir a um(a) usuário(a) construir ferramentas para ter diálogos específicos com leitores e leitoras de sua interpretação. Este tipo de tecnologia já foi investigado por Monteiro (Monteiro, 2015; Monteiro *et al.*, 2013a,b) e aparentemente aponta para caminhos muito promissores na via da hermenêutica digital.

O tema de longo prazo que me interessa, a influência de novas formas de "leitura" de texto na produção de conhecimento, ganhou com este estudo motivação renovada. São muitas as possibilidades e os desafios, dentre os quais alguns riscos para a validade científica de pesquisa informatizada. Penso especificamente no que vem a ser uma pesquisa qualitativa bem feita, neste novo cenário tecnológico. Parte desta questão é tratada por Romele em seus trabalhos já citados, mas há bem mais a investigar. De imediato, ainda em uma linha modesta de investigação, com o objetivo de aprofundar minha competência técnica com o Voyant e enriquecer o contexto de interpretação, há três linhas a seguir. A primeira é a de trabalhar com textos muito mais volumosos; a segunda, a de trabalhar com mais intérpretes participantes; e a terceira, a de explorar a *programabilidade* do Voyant. Tendo feito isto, acredito que estarei em bem melhor condição para estudar o que gostaria de saber sobre novas condições de produção (e validação?) de conhecimento.

Referências Bibliográficas

1. Drucker, J. (2017). Why Distant Reading Isn't. *Pmla*, 132(3), pp. 628–635. <https://www.jstor.org/stable/27037376>
2. Monteiro, I. T. (2015). *Autoexpressão e engenharia semiótica do usuário-designer*. Rio de Janeiro, 2015. 312p. Tese de Doutorado. Orientadora: Clarisse Sieckenius de Souza. Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro. <https://repositorio.ufrj.br/handle/riufc/13306>
3. Monteiro, I. T., Tolmasquim, E. T. e de Souza, C. S. (2013a) **Going back and forth in metacommunication threads**. In *Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems (IHC '13)*. Brazilian Computer Society, Porto Alegre, BRA, 102–111. <http://www3.serg.inf.puc-rio.br/docs/IHC2013-Monteiro-Tolmasquim-deSouza.pdf>
4. Monteiro, I. T., de Souza, C.S. e Leitão, C. S. (2013b). **Metacommunication and semiotic engineering: insights from a study with mediated HCI**. In *Proceedings of the Second international conference on Design, User Experience, and Usability: design philosophy, methods, and tools - Volume Part I (DUXU'13)*, Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 115–124. https://doi.org/10.1007/978-3-642-39229-0_14
5. Moretti, F. (2007). *Graphs, maps, trees*. Verso, New York.
6. Moretti, F. (2013). *Distant Reading*. Verso, New York.
7. Moretti, F. (2017) "Patterns and Interpretation." *Stanford Literary Lab: Pamphlets ; 15*. Pamphlets of the Stanford Literary Lab, 2017. <https://litlab.stanford.edu/LiteraryLabPamphlet15.pdf>
8. OpenAI (2023). "GPT-4 Technical Report," March 2023. <https://doi.org/10.48550/ARXIV.2303.08774>.
9. Papa Francisco (2024). Mensagem do Santo Padre Francisco para a Celebração do Dia Mundial da Paz, 1 de Janeiro de 2024. Online em: <https://www.vatican.va/content/francesco/pt/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html> Última visita em Abril de 2024.
10. Romele, A., Severo, M. and Fúria, P. (2020). "Digital hermeneutics: from interpreting with machines to interpretational machines," *AI & SOCIETY*, vol. 35, no. 1, pp. 73–86, doi: 10.1007/s00146-018-0856-2.
11. Romele, A. (2021). *Digital hermeneutics*. London. Routledge.
12. Rockwell, G. e Sinclair, S. (2016). *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: The MIT Press.
13. Santaella, L. (2019). O livro como prótese cognitiva. *MATRIZES*, 13(3), pp. 21–35. DOI: 10.11606/issn.1982-8160.v13i3p21-35
14. The Peirce Edition Project (1998) *The Essential Peirce. Selected Philosophical Writings. Volume 2 (1983-1913)*. Bloomington, IN. Indiana University Pres..