Criptografia

O código Voynich

Físicos brasileiros propõem significado para termos de manuscrito produzido em linguagem desconhecida

Igor Zolnerkevic

O mais enigmático dos livros que se conhece – o manuscrito Voynich, um texto supostamente do início do século XV composto em um alfabeto desconhecido – parece não ser um amontoado aleatório de símbolos sem sentido, como afirmam alguns estudiosos. Ao menos, essa é a conclusão a que chegou um grupo de físicos brasileiros depois de usar técnicas estatísticas avançadas para analisar esse documento que há tempos frustra os maiores especialistas em decifrar códigos criptografados.

Pouco se sabe sobre o manuscrito e sua história. Apenas que está escrito em um alfabeto inventado, jamais visto em outro documento, e que foi adquirido em 1912, perto de Roma, na Itália, por um livreiro polonês chamado Wilfried Voynich, que se casou com a filha de George Boole, famoso matemático britânico. O manuscrito é ricamente ilustrado com imagens de plantas e corpos celestes, o que sugere que se trate de um texto sobre ervas e astrologia. Mas seu conteúdo continua um enigma.

Os físicos também não o decifraram, mas, em sua análise, publicada em julho na revista *Plos One*, acreditam ter identificado suas palavras-chave, isto é, o conjunto das palavras que mais se aproximam dos nomes dos tópicos abordados no texto. Uma futura tradução dessas palavras poderia finalmente revelar algo da mensagem do livro, se é que de fato existe uma mensagem. A equipe brasileira, formada por pesquisadores que atuam na Alemanha e em São Carlos, interior de São Paulo,

também concluiu que o texto do manuscrito apresenta todas as propriedades estatísticas que se espera de um texto com significado. Se eles estiverem certos, o manuscrito não seria uma sequência de símbolos sem sentido.

De qualquer forma, o método desenvolvido pelos pesquisadores para estudar o Voynich tem outras aplicações. “Ele nos permite identificar as palavras-chave de um texto longo sem que seja necessário conhecer sua organização ou compará-lo com outros textos, tal como fazem mecanismos de busca como o do Google”, explica um dos autores do estudo, o físico Eduardo Altmann, do Instituto Max Planck para Física de Sistemas Complexos, em Dresden, Alemanha. Altmann vem discutindo com funcionários da biblioteca de Dresden o uso de um sistema de classificação automática de documentos que encontraria palavras potencialmente importantes, que tenham passado despercebidas durante a classificação dos livros pelos bibliotecários. “Esse sistema poderia ajudar a encontrar conexões entre disciplinas científicas diferentes”, diz.

Altmann vem se especializando desde 2009 em usar técnicas da física estatística para analisar a frequência com que as palavras aparecem ao longo de um texto (*ver Pesquisa FAPESP nº 185*). Durante uma conferência, conheceu outro brasileiro radicado na Alemanha, o físico Diego Rybski, do Instituto para Pesquisa de Impacto no Clima, em Potsdam, que conhecia a história do manuscrito Voynich.

Fraude ou não?

Atualmente guardado na Biblioteca Beinecke da Universidade de Yale, Estados Unidos, o manuscrito se encontrava perdido no meio de uma coleção mantida por padres jesuítas italianos, quando foi adquirido por Voynich. Especulou-se muito se o manuscrito não seria uma fraude criada pelo próprio Voynich, que lucrou com sua venda, mas historiadores e biógrafos já descartaram essa hipótese.

Junto ao manuscrito, uma carta datada de 1666, assinada por um acadêmico da cidade de Praga, na atual República Tcheca, pedia a um jesuíta em Roma que tentasse

decifrá-lo. A correspondência menciona que o manuscrito pertenceu a Rodolfo II (1552-1612), imperador do Sacro Império Romano-Germânico, conhecido por seu fascínio pelo ocultismo, e que o autor do livro talvez fosse o filósofo e frade franciscano inglês Roger Bacon, que viveu de 1214 a1294. Uma análise físico-química dos papéis e das tintas feita em 2010, contudo, concluiu que o manuscrito deve ter sido produzido entre 1404 e 1438.

Com dimensões de um livro de bolso e encadernadas em papel velino, as 240 páginas do manuscrito são ricamente ilustradas, sendo que algumas folhas têm de duas a três vezes o tamanho do livro quando desdobradas. A temática dos desenhos é a única pista sobre de que tratam as suas seções. Metade do volume retrata plantas inteiras, a maioria não identificada (três delas foram, mas as espécies ocorrem em várias partes do mundo, não ajudando a localizar sua origem). Segue uma seção astrológica, com desenhos do Sol, da Lua, de estrelas, o zodíaco, círculos no céu e muitas mulheres nuas. A seção seguinte contém estranhos desenhos de tubos, que se conjectura serem vasos sanguíneos, microscópios ou telescópios, e mais mulheres nuas em piscinas. Em seguida vem a seção chamada de farmacêutica, que parece uma lista aparentemente sobre folhas e raízes. O livro termina com páginas repletas de um texto formado por uma série de parágrafos curtos, ilustrado apenas por estrelas nas margens.

Os cerca de 40 símbolos do texto lembram vagamente números arábicos e letras do alfabeto latino, bem como alguns sinais usados por alquimistas medievais. Eles estão organizados como em qualquer texto ocidental, agrupados em palavras separadas por espaços. Ao mesmo tempo familiar e única, a linguagem do manuscrito ludibriou todos os especialistas que a examinaram. Foi uma obsessão para o criptoanalista norte-americano William Friedman, famoso por decifrar códigos secretos dos alemães e dos japoneses durante a Segunda Guerra Mundial. Após 20 anos de tentativa, Friedman chegou apenas ao palpite de que sua mensagem estava escrita em uma língua inventada.

A partir dos anos 1990, uma comunidade formada por uma centena de pesquisadores de várias disciplinas interessados no Voynich começou a se comunicar pela internet. Para facilitar a discussão de partes do texto via email, associaram cada “letra” da escrita voynichesa a um caractere latino. Essa transcrição facilitou a análise estatística do texto por computador e a sua comparação com outros textos.

Uma dessas análises ganhou repercussão em 2004, sugerindo que o Voynich seria uma fraude. O psicólogo e matemático Gordon Rugg, da Universidade Keele, na Inglaterra, descobriu como criar uma sequência de símbolos semelhante à escrita voynichesa por meio de técnicas criptográficas disponíveis no período renascentista. Rugg acredita que o livro seja obra de um charlatão do século XVI, de olho no ouro que Rodolfo II oferecia por relíquias místicas. Em 2007, o físico Andreas Schinner, da Universidade Johannes Kepler, na Áustria, sugeriu em análise publicada na revista *Cryptologia* que o texto do Voynich teria sido criado por um processo aleatório. Esses resultados, porém, não desanimaram a maioria dos estudiosos do Voynich, nem abalaram a fé de místicos que acreditam que o manuscrito contenha alguma profecia divina ou alienígena.

“O tamanho da literatura sobre o Voynich é assustadora e me fez perguntar até que ponto seu objetivo é científico”, conta Altmann. “É por isso que em nosso trabalho tentamos formular as questões de maneira geral, esperando que o estudo tenha outras aplicações.”

Rede de palavras

Altmann e Rybski colaboraram com os físicos Osvaldo Oliveira Jr. e Luciano da Fontoura Costa, do Instituto de Física de São Carlos, da Universidade de São Paulo, que tratam textos como se fossem redes complexas de palavras (*ver figura na página XX*). “Duas palavras são conectadas na rede se elas aparecem vizinhas no texto”, explica Diego Amâncio, aluno de doutorado de Costa e primeiro autor do artigo.

Antes de atacarem o Voynich, os pesquisadores avaliaram 29 tipos de medidas

estatísticas diferentes que podem ser obtidas da análise de um texto qualquer. Elas são quantidades que medem como as palavras se aglomeram ou se dispersam ao longo do texto ou que medem a distribuição dos vários arranjos possíveis das conexões entre as palavras, quando o texto é representado como uma rede complexa.

“Fazendo um experimento imaginário, se pudéssemos analisar todos os textos já escritos em todas as línguas existentes teríamos todos os valores possíveis para essas medidas estatísticas”, explica Altmann. Nessa situação ideal, os pesquisadores teriam então uma espécie de assinatura estatística para cada texto possível.

Altmann procurou textos longos, traduzidos em mais de dez línguas, que estivessem digitalizados e sem restrições de direitos autorais, para análise. Encontrou apenas um, o Novo Testamento (Eduardo, esta informação está correta? Há vários clássicos da literatura traduzidos para muitas línguas). A equipe analisou o texto bíblico em quinze línguas, do árabe ao xhosa, falada por um pequeno grupo da África do Sul,e comparou com versões do texto com as palavras embaralhadas. Conseguiram assim ter uma ideia de quais medidas estatísticas são mais sensíveis a variações na língua e quais delas podem distinguir um texto com significado de uma série aleatória de palavras.

Os pesquisadores compararam também vários textos escritos em uma mesma língua, analisando quinze romances clássicos da literatura portuguesa e quinze da inglesa, assim como suas versões embaralhadas, determinando quais das medidas dependem mais da mensagem particular do texto do que da língua em que está escrito.

Ao contrário da análise de Schinner, as medidas avaliadas pelos brasileiros indicam que o texto do Voynich tem estrutura sintática e transmite alguma mensagem. “Na minha experiência, os resultados do Schinner não são necessariamente uma indicação de que o texto não esteja escrito em uma língua natural”, diz Altmann. Ele explica ainda que, se tivessem comparado com mais livros, seria possível dizer até qual língua é a mais próxima da usada no manuscrito.

Foi combinando algumas dessas medidas, que os pesquisadores desenvolveram um método para distinguir palavras do manuscrito cujo significado remete aos tópicos abordados no texto em geral das palavras que cumprem papel meramente sintático, como artigos e preposições. Aplicado ao Novo Testamento em português, por exemplo, o método resulta em uma lista de palavras incluindo “Maria”, “nasceu”, “menino”, “sepulcro” e “bem-aventurados”. A lista do Voynich inclui “cthy”, “qokeedy” e “shedy”.

Um estudo publicado uma semana antes na mesma revista chegou a conclusões parecidas. Os físicos argentinos Marcelo Montemurro, da Universidade de Manchester, na Inglaterra, e Damián Zanette, do Instituto Balseiro, na Argentina, usaram técnicas estatísticas diferentes e chegaram a uma lista muito similar de palavras importantes.

“Cada novo estudo sobre o manuscrito aponta detalhes que são característicos de línguas naturais e improváveis em textos aleatórios”, observa Jorge Stolfi, do Instituto de Computação da Universidade Estadual de Campinas (Unicamp), que analisou o manuscrito entre 1997 e 2004. “Arrisco dizer que é uma transcrição fonética de alguma língua do leste asiático, feita por um europeu, provavelmente ditado por um nativo, usando um alfabeto inventado pelo autor para esse fim.”

“Infelizmente, não sei como progredir nessa direção”, diz Stolfi, cuja teoria causou certo *frisson* ao ser apresentada no ano passado em uma conferência comemorando o centenário da descoberta do manuscrito. “Mesmo que minha hipótese esteja correta, não me arrisco a prever quando será decifrado.”

Artigo científico

AMANCIO, D. R. *et al*. Probing the Statistical Properties of Unknown Texts: Application to the Voynich Manuscript. Plos One. v 8(7). Jul. 2013.

PROJETOS

1. Uso de Redes Complexas no Processamento de Línguas Naturais (nº 2010/00927-9); **Modalidade** Doutorado Direto; **Benefic.** Diego Rafael Amancio; **Investimento** R$ 109.708,56 (FAPESP); 2. Modelos e métodos de *eScience* para ciências da vida e agrárias (nº 2011/50761-2); Modalidade Projeto Temático; Coord. Roberto Marcondes Cesar Junior / IME-USP; **Investimento** R$ 1.033.785,69 (FAPESP/CNPq)