

**CENTRO UNIVERSITÁRIO BELAS ARTES DE SÃO PAULO
ARTES VISUAIS**

**ARTE CONTEMPORÂNEA A PARTIR DO PRÊMIO PIPA:
UM MODELO DE APRENDIZADO DE MÁQUINA
VOLTADO A CRÍTICA DE ARTE**

**Orientando: Fábio Hideki Eguchi
Orientador: Prof. Dr. Cauê Alves**

RESUMO:

Este artigo apresenta um modelo de compreensão de linguagem natural baseado em textos críticos de arte contemporânea brasileira. Desenvolvido utilizando serviços de inteligência artificial *Watson* da *IBM Cloud*, o modelo foi treinado com técnicas de aprendizado de máquina supervisionado usando um conjunto de textos críticos sobre artistas que já foram finalistas do Prêmio PIPA. Implementado em um aplicativo web como prova de conceito, em seu estágio de desenvolvimento inicial ele é capaz de identificar em um texto pessoas, obras, técnicas e extrair suas respectivas relações. Além disto, foram definidas diretrizes para a continuidade do desenvolvimento do modelo, expandindo sua capacidade de identificação de entidades mencionadas e extração de relações.

PALAVRAS-CHAVE: Crítica de Arte. Processamento de linguagem natural. Arte contemporânea brasileira. Aprendizado de máquina.

ABSTRACT:

This article presents a natural language understanding model based in contemporary Brazilian art critiques. The model was developed using Watson's artificial intelligence services from IBM Cloud, and trained with supervised machine learning techniques using a set of critiques from PIPA Prize finalist's artists. Implemented on a web application as a proof of concept, the model at an early development state can recognize people, artworks, techniques and extract its relations in an art related text. Furthermore, guidelines were set for future model development, increasing its named entity and recognition and relation extraction capabilities.

KEYWORDS: Art criticism. Natural language processing. Contemporary Brazilian art. Machine learning.

1. INTRODUÇÃO

O interesse por processar a linguagem humana surge junto com a própria computação ao final da Segunda Guerra Mundial¹. Tópico de estudo de diversas áreas do conhecimento, em ciências da computação tal campo é cunhado como *Processamento de linguagem natural* (NLP). Presente no nosso cotidiano mais notadamente nos sites de buscas, corretores ortográficos e assistentes virtuais, tal campo teve crescimento expressivo no final da década de

¹ Um acontecimento precursor disto é o modelo de computação algorítmica de Alan Turing em 1936, que levou a Noam Chomsky em 1956 considerar Máquinas de Estados Finitos como um modo de caracterizar a gramática. Tal período é tido como a base tanto da ciência da computação, quanto da linguística moderna através da escola de Chomsky. (JURAFSKY; MARTIN, 2008, p. 9).

noventa com o advento da internet e consequente melhoria dos computadores para uso pessoal e comercial.

Já no início dos anos 2000, a disponibilização de uma grande quantidade de materiais, escritos e falados, por grandes consórcios da área permitiram a exploração mais complexa de ferramentas de análise e interpretação semântica com técnicas de aprendizado de máquina supervisionado². Neste contexto, atualmente as maiores plataformas de serviços de computação oferecem algum tipo de serviço que alie NLP à aprendizado de máquina³. As funções mais comuns neste tipo de serviço é a categorização/classificação de documentos, análise de sentimento/emoção, identificação de palavras/frases chave, extração de relações (*relation extraction*) e identificação de entidades mencionadas (*named entity recognition*) como pessoa, data e lugar. Tendo em vista o nível de complexidade da linguagem humana, e as particularidades de um léxico, essas plataformas oferecem serviços que possibilitam a customização de um modelo de NLP para uma área específica do conhecimento, possibilitando que a função de identificação de entidade passe a reconhecer novos tipos de entidades. O *Comprehend Medical* da *Amazon Web Services* é um desses serviços que utiliza um modelo de NLP voltado para à área da saúde, e que permite identificar e relacionar em um texto não-estruturado⁴ medicamentos, sintomas, exames e tratamentos através de aprendizado de máquina. Este podendo ser implementado em aplicações que ajudem em decisões clínicas, ou em um sistema de triagem, tendo em vista clientes como empresas farmacêuticas, redes de hospitais e centros de pesquisa da área da saúde.

Neste contexto, este artigo apresenta o desenvolvimento de um modelo de NLP baseado em aprendizado de máquina e especializado em textos críticos do âmbito da arte contemporânea brasileira. Criado na plataforma de inteligência artificial *Watson* da *International Business Machines Corporation* (IBM), o modelo em seu estado atual de desenvolvimento consegue identificar em um texto as pessoas, obras e técnicas artísticas citadas; e extrair suas respectivas relações. Além disto, foi estabelecido um sistema de tipo com mais tipos de entidades e relações, servindo assim como uma base para modelos mais complexos. Ciente de ser uma pesquisa introdutória no campo de pesquisa sobre arte contemporânea em relação ao

² Idem, ibidem, p. 12.

³ Como o *AutoML Natural Language* do *Google Cloud*, a *API de Análise de Texto* do *Microsoft Azure* e o *Amazon Comprehend* da *Amazon Web Services*.

⁴ Neste âmbito um texto não estruturado é uma extensão da noção de dados estruturados e não estruturados. Podemos associar tal conceito a um formulário de pesquisa. Os campos de respostas com restrições claras como Nome, e-mail, data de nascimento e perguntas quantitativas geram dados estruturados, aqueles que já estão em um formato próprio para um banco de dados. Já um campo de pergunta qualitativa, aquela que dá espaço para uma resposta aberta, geraria um dado não estruturado. O desafio do NLU é justamente fazer tal estruturação sem perder a complexidade de informações que neste caso uma resposta aberta pode trazer.

processamento de linguagem natural, o presente artigo se propõe também a discutir aspectos elementares da arte sob uma ótica atual. Tendo em vista que o uso de inteligência artificial⁵, ou até mesmo ciência de dados⁶, no campo de pesquisa sobre artes visuais ainda é emergente em comparação com outros campos do conhecimento. Há também, uma tentativa de não negligenciar os aspectos sociais inerentes a tal tipo de tecnologia, ao pontuar possíveis implementações de tal modelo em aplicações que democratizem o acesso a arte e estimulem a crítica e pesquisa.

2. FLUXO DE CRIAÇÃO DO MODELO

O serviço *Watson Knowledge Studio* (WKS) da *IBM Cloud* foi escolhido para o desenvolvimento do modelo por este ter uma versão gratuita, estar em um estado de desenvolvimento maduro no momento de início desta pesquisa, fornecer suporte para treinar o modelo em português do brasil e permitir integração com outros serviços que possibilitem implementar e testar o modelo em aplicativos de código aberto.

Nesta plataforma, o fluxo de criação do modelo (fig.1) pode ser simplificado em três etapas: adição de recursos, anotação dos documentos e por último treinamento e avaliação do modelo (IBM..., 2019a). Na etapa de adição de recursos são definidas as bases do modelo, quais documentos servirão como fonte de aprendizado e quais são os elementos de interesse destes documentos através de um sistema de tipos. Os documentos são divididos em três grupos: treinamento, teste e cego. A etapa de anotação é a de ensino, onde o sistema de tipos é aplicado ao grupo de documentos de treinamento através de anotações manuais⁷. Feito isto é realizado o treinamento do modelo ao enviar os conjuntos de documentos citados anteriormente, que após algumas horas ou minutos retorna uma versão pronta para o uso. Além disso é gerada uma avaliação, ao aplicar o modelo treinado aos grupos de teste e cego, que detalha a precisão do modelo e indica os pontos com baixa precisão. Após isso, as etapas anteriores podem ser revistas, novos documentos podem ser incluídos, o sistema de tipos pode ser ajustado, e novas anotações realizadas, para então ser feito um novo treinamento e criar uma versão nova do modelo.

⁵ Nos últimos anos projetos como *A Voz da Arte*, realizado em 2017 na Pinacoteca de São Paulo, e a *Outra 33ª Bienal de São Paulo*, realizado em 2018, foram emblemáticos neste campo.

⁶ É o caso do projeto *A HISTÓRIA DA _RTE*, apoiado pelo Rumos Itaú Cultural 2015-2016.

⁷ É tal procedimento que dá o caráter de *supervisionado* ao aprendizado de máquina. No aprendizado supervisionado os dados de entrada, no nosso caso as anotações, são associados a um dado de saída correto (um “sinal do supervisor”). O objetivo do algoritmo é aprender a mapear a partir de uma nova anotação um dado de saída correto. JURAFSKY, Daniel; MARTIN, James. op. cit., p. 57.

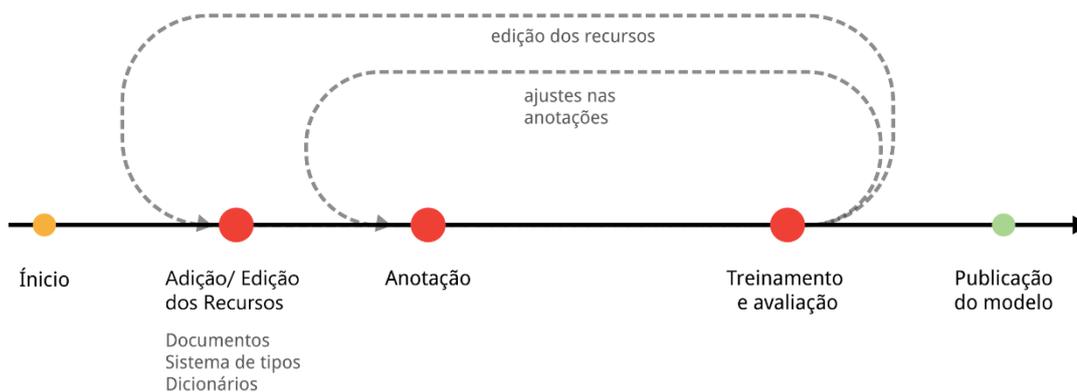


Figura 1 – Fluxograma de criação do modelo

Fonte: Autoria própria

2.1 - CRÍTICA E ARTE BRASILEIRA ATUALMENTE

O texto crítico foi elencado como fonte para o conjunto de documentos utilizados no treinamento do modelo, tendo em vista que este, em sua maioria, apresenta mais conteúdo em relação à um texto de jornalismo cultural e é mais recorrente que os escritos de artista. Neste âmbito, o lugar da crítica de arte no Brasil passa por profundas transformações a partir da década de 80. Com a crescente diminuição em seu espaço de veiculação na mídia impressa em detrimento do jornalismo cultural, há uma migração para plataformas mais específicas e ligados ao próprio sistema da arte. Tal situação é essencial para compreender a acepção a atual do *texto crítico*, que hora pode possuir vínculos maiores com o âmbito acadêmico, e hora da curadoria (AMADO, 2006, p. 243). Sendo assim, um texto crítico atualmente é um termo que abarca também texto para exposições; principalmente o curatorial presente na parede de entrada na maioria das exposições, textos feitos para publicações como catálogos e livros de trajetória de artistas; e textos ensaísticos que provém do âmbito acadêmico.

Para além de tal cenário regional, é importante ressaltar também as mudanças da arte moderna para a contemporânea de maneira geral, e por conseguinte, na produção textual acerca desta. A arte contemporânea dilui não só os limites entre suas linguagens tradicionais, mas também para com o mundo. A arte começa a falar e se apropriar de várias regiões do saber e do cotidiano. Cocchiarale (2006, p. 219) afirma que o curador deve “produzir questões, quase sempre extraestéticas, temáticas, que emprestem sentido, ainda que provisório, à dispersão aparente em que nos encontramos”. Sendo assim, a compreensão de um texto crítico não fica restrita ao campo específico da arte, mas se abre para uma gama imprevisível de assuntos⁸.

⁸ Por exemplo, um texto crítico acerca de uma obra sítio-específico em geral fornece um contexto acerca do local em que tal obra está exposta. Durante a pesquisa notamos que tal contexto pode ter um viés mais histórico, geográfico, sócio-político ou arquitetônico.

Tendo em vista essa amplitude da arte contemporânea, e que um dos objetivos da presente pesquisa não é elencar artistas; para a elaboração do modelo capaz de compreender um texto crítico usamos como base os artistas participantes do Prêmio PIPA. Com 10 anos de trajetória, a cada edição anual, o Conselho do PIPA convida de 20 a 40 profissionais do sistema da arte brasileiro para indicar à premiação artistas com destaque no circuito nacional. “Os convidados são profissionais renomados que atuam com arte contemporânea, como artistas com carreira com trajetória reconhecida, críticos da arte, colecionadores, galeristas e curadores” (Instituto PIPA, 2019, p. 6). Tal prêmio tem como objetivo “consagrar artistas já conhecidos no circuito de arte brasileiro que vêm se destacando por seus trabalhos” (Instituto PIPA, 2019, p. 6).

2.2 – CONJUNTO DE DOCUMENTOS

Partimos de um universo inicial de 430 artistas participantes das 10 primeiras edições do Prêmio PIPA, entre 2010 e 2019. Cada artista possui um perfil no site do prêmio contendo imagens de seus trabalhos, links externos para seus sites pessoais e de representação comercial, e eventualmente textos críticos. Após um levantamento geral da disponibilidade de textos nestas fontes, reduzimos nosso universo para um grupo de 51 artistas finalistas ou vencedores e vencedoras em alguma categoria do Prêmio⁹. Em seguida, criamos um banco de dados e filtramos quais finalistas possuíam textos críticos disponíveis neste perfil, ou em seções dos sites mencionados anteriormente. Tendo em vistas as diretrizes recomendadas para os documentos no WKS (IBM..., 2019c), e a quantidade média de textos por artista que foi observada, utilizamos como critério para integrar o conjunto de documentos a disponibilidade de 3 ou mais textos¹⁰ de um mesmo artista nos sites citados. Resultando assim na amostra de 9 artistas (tabela 1) e 30 textos.

Esta quantidade de textos de um mesmo artista possibilitou que menções semelhantes fossem inclusas nos grupos de treinamento, teste e cego de maneira mais consistente. O corpo de cada texto crítico foi copiado para um arquivo do tipo DOC e nomeado com um código alfanumérico de controle, enquanto suas informações adicionais foram registradas em um banco de dados. Além disso, foram suprimidas epígrafes e notas de rodapé de cada texto, tendo

⁹ Tal informação consta no perfil de tais artistas no site do Prêmio PIPA. Na categoria principal o vencedor do Prêmio PIPA é escolhido pelo júri de premiação dentre uma lista de 4 finalistas indicados pelo Conselho do PIPA. Outra categoria elege um vencedor destes 4 finalistas a partir de uma votação aberta ao público de uma exposição realizada com eles. Já a categoria PIPA Online, chamada de PIPA Online Popular em 2014 e 2015, é aberta a todos os artistas indicados da edição, sendo o vencedor ou vencedora a pessoa que recebeu mais votos online.

¹⁰ Nesta etapa também foi avaliada a relevância de cada texto para a pesquisa, tendo em vista os critérios de média de 1000 palavras por textos e quantidade de menções relevantes, ou seja, aquelas com entidades e relações do sistema de tipos aplicado.

em vista o fluxo de anotação. Sendo assim, o conjunto de documentos utilizado para o treinamento do modelo conteve o total de 21 590 palavras. É importante notar que a plataforma recomenda um conjunto de cerca de 300 000 palavras como meta para um modelo robusto, variando de acordo com o nível de complexidade desejado (IBM..., 2019c). Neste sentido, o conjunto utilizado serviu para os estágios iniciais do modelo, de aferir a viabilidade do sistema de tipos e realizar o treinamento das entidades e relações principais. Portanto, a expansão do conjunto de documentos é necessária para continuidade do desenvolvimento do modelo em oportunidades futuras.

Artista	Local de nascimento	Local de atuação	Fonte textos
Ana Paula Oliveira	Uberaba, MG	São Paulo, SP	Site da Artista
André Komatsu	São Paulo, SP	São Paulo, SP	Site da Galeria
Cabelo	Cachoeira de Itapemirim, ES	Rio de Janeiro, RJ	Site do Artista
Carla Guagliardi	Rio de Janeiro, RJ	Rio de Janeiro, RJ e Berlim, Alemanha	Site do PIPA
Cynthia Marcelle	Belo Horizonte, MG	Belo Horizonte, MG	Site da Galeria
Eduardo Berliner	Rio de Janeiro, RJ	Rio de Janeiro, RJ	Site da Galeria
Luiza Baldan	Rio de Janeiro, RJ	Santiago, Chile	Site da Artista
Marcelo Moscheta	São José do Rio Preto, SP	Campinas, SP	Site do Artista
Tatiana Blass	São Paulo, SP	São Paulo, SP	Site da Artista

Tabela 1 – Relação dos artistas da amostra. Fonte: Autoria própria.

2.3 - SISTEMA DE TIPOS

Na plataforma WKS um sistema de tipos define o conteúdo de interesse do modelo desenvolvido. É possível criar tipos de entidades e tipos de relação, permitindo assim anotar menções nos textos com tais tipos. Por exemplo, a menção *Tatiana Blass* é anotada com o tipo de entidade *PESSOA*. A menção *Zona Morta* é anotada com o tipo de entidade *OBRA*. Depois que uma menção é anotada com um tipo de entidade, o período anotado de texto é referido como uma entidade. Por exemplo, a entidade *Tatiana Blass* pode ser mencionada durante o texto diversas vezes, seja em menções escritas de maneira idêntica ou de outras formas como *Blass*, *Tatiana*, *a artista*, *ela*, *sua*, *dela*. Porém, todas essas menções se referem a mesma entidade *Tatiana Blass* que foi anotada com o tipo de entidade *PESSOA*. Já os tipos de relação definem um relacionamento ordenado e binário entre duas entidades. Uma entidade pode ter várias relações com outras entidades. Por exemplo, dentro de uma sentença é anotado um tipo de relação *obraDe* entre a entidade *Tatiana Blass* e a entidade *Zona Morta*, e a relação *atributoDe* entre a entidade *Zona Morta* e *instalação*.

No início, da pesquisa foram mapeadas diversas entidades e relações de interesse, pensando de maneira ampla o circuito da arte, seus agentes e relações. Tendo *ARTISTA*¹¹ e *OBRA* como tipos de entidade centrais que interagem com as demais entidades (fig.2). Após o contato com o conjunto de documentos elencado acima, e testes na plataforma WKS, o principal motivo para descartar tais tipos de entidades periféricas foi a baixa quantidade de menções destes, e conseqüentemente de relações com estes. Como por exemplo *EVENTO*, *LOCALIZAÇÃO* e até mesmo *DATA*. Neste último caso, o tipo de relação *criadaEm* para os tipos de entidade *OBRA* e *DATA* apresentou um número baixo de menções nas fases iniciais da pesquisa. Além disto, um fator percebido após um contato maior com o conjunto de documentos é que há uma certa higienização do texto em relação ao seu contexto. De maneira geral, há um ar de atemporalidade no conteúdo escrito, em que a situação sempre envolve um público visitante implícito, nós, diante de uma exposição oculta. Raros são os casos em que o autor contextualiza de maneira direta a obra descrita. Alguns chegam a colocar observações em notas de rodapé, ou em uma nota final de maneira sucinta, que tal texto foi escrito em ocasião de uma exposição ou catálogo, e o espaço expositivo onde se encontram as obras referidas

Dada as características do conjunto de documentos e as qualidades esperadas do modelo¹², definimos então um sistema de tipos inicial tendo como centro o tipo de entidade *OBRA* (fig.3). Ao longo das etapas de definição das diretrizes de anotações e teste de viabilidade do sistema de tipos, também houve ajustes decorrentes dos resultados obtidos e das questões que observamos ao decorrer deste processo. Como por exemplo, a criação do tipo de entidade *SENTIDO* e a junção dos tipos de entidade *MATERIA* e *CARACTERISTICA* por desempenharem papel semelhante nos textos. Além da ampliação de tipos de relações atribuídos somente a *OBRA* para *PESSOA* e vice-versa como por exemplo a relação *faz* com *PROCEDIMENTO*. Durante a etapa de criação das diretrizes de anotação, tal sistema de tipos se mostrou capaz de cobrir grande parte do conteúdo dos textos. Por conta do nível de complexidade encontrado nos textos e duração da pesquisa, o modelo chegou em um nível de desenvolvimento inicial. Foram treinados os seguintes tipos de entidades e suas respectivas

¹¹ Outros tipos de entidades referentes a pessoas e que integram o sistema da arte também foram pensados, porém, a menção destes foram pouco frequentes no conjunto de documentos. Como curadores, galeristas e arquitetos. Uma opção para envolver essas menções seria utilizar o tipo de entidade *jobTitle*, presente no sistema de tipos padrão do NLU. Porém optamos por mudar o tipo de entidade *ARTISTA* para *PESSOA*, abrangendo assim menções também como *público* e *visitantes*. Em um sistema mais complexo, ambas as entidades poderiam ser utilizadas, tendo em vista a possibilidade de rotular a função de uma menção como citado na seção 2.5.

¹² O caráter crítico provido pelo conjunto de documentos foi um fator balizar para o sistema de tipos elaborado. Para treinar um modelo que identificasse a data de execução de uma obra, e o local que ela foi apresentada, provavelmente seria mais eficiente usar textos de legendas em sites e catálogos no conjunto de documentos e um modelo baseado em regra, e não aprendizado de máquina.

relações: *PESSOA*, *OBRA*, *TECNICA*, *obraDe*, *atributoDe*. o espaço expositivo onde se encontram as obras referidas.

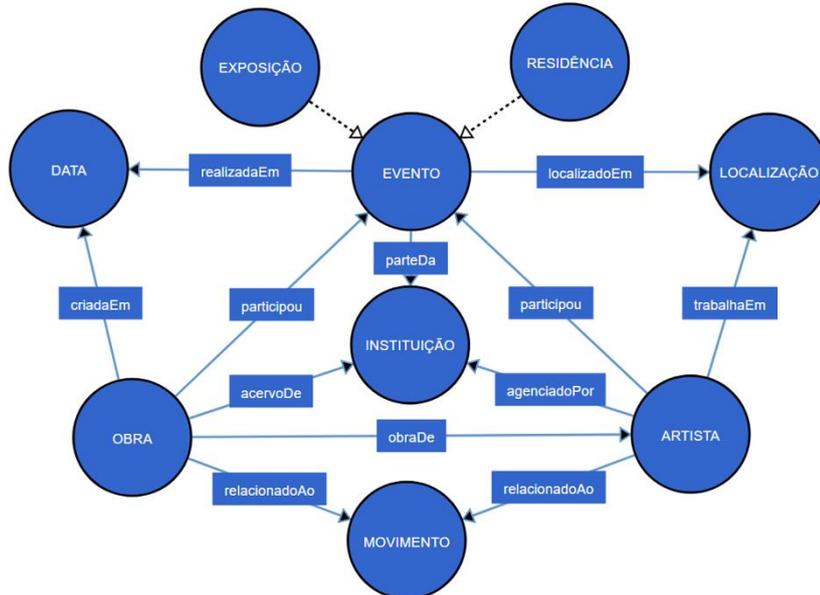


Figura 2 – Organograma do sistema de tipos mapeado inicialmente.

Legenda: Os círculos representam as entidades, as flechas simbolizam as relações entre as entidades, os retângulos os tipos de relações e a linha tracejada indica que é um subtipo de entidade. Fonte: Autoria própria.

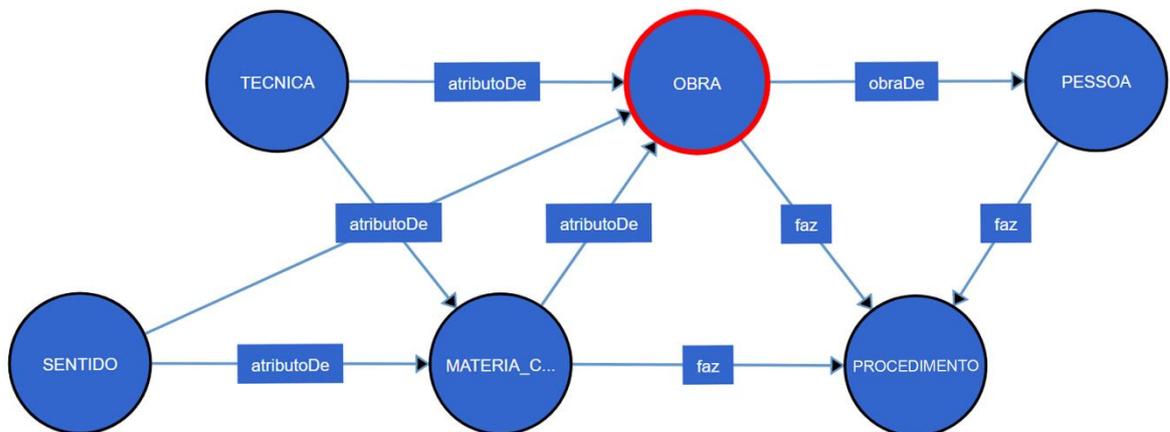


Figura 3 – Organograma do sistema de tipos elaborado.

Legenda: Os círculos representam as entidades, as flechas simbolizam as relações entre as entidades e os retângulos os tipos de relações. Fonte: Autoria própria.

2.4 – DIRETRIZES DE ANOTAÇÃO

No ambiente de anotação do WKS cada documento é acessado individualmente, e suas sentenças são numeradas e ordenadas na interface. O fluxo de anotação começa no modo *Menção*, onde as entidades são rotuladas com seus tipos de entidades respectivos. Tanto o modo *Relação*, quanto o modo *Correferência* que são opcionais foram utilizados em todos os documentos, tendo em vista que desejávamos que o modelo identifique relações. E que menções dos tipos de entidade *PESSOA* e *OBRA* são comumente repetidas ao longo de um documento.

Neste contexto é importante ressaltar uma limitação técnica existente na plataforma: Só é possível rotular relações entre entidades de uma mesma sentença¹³. Porém, no modo *Correferência* é possível indicar que várias menções distintas correspondem a mesma entidade no texto. Ainda assim, textos que utilizem sentenças com sujeito oculto não são bem anotados, tendo em vista que algumas menções de entidades acabam não sendo rotulados, e por consequência, as relações que essas entidades estabelecem também não. É o caso do exemplo da figura 4. *Moscheta*, o artista objeto da crítica é mencionado na sentença 11 e rotulado com o tipo de entidade *PESSOA*. Já na sentença seguinte, por possuir um sujeito oculto não é possível rotular relações entre *Moscheta* e as entidades *reproduziria* e *série Carbon copy*. Entretanto, na sentença 13, a *TECNICA gravura em metal* pode ser relacionada ao *fundo* da *OBRA* que é mencionada novamente de forma nominal, através da palavra *trabalho* (fig.5). Ao criar uma cadeia de coreferências entre estas menções, o tipo de relação *atributoDe* é estabelecido de maneira correta, mantendo assim a coerência na análise do texto.

Entity		Mention
Type	Subtype	Role
4	MATERIA_CHARACTER	
2	OBRA	
1	PESSOA	
6	PROCEDIMENTO	
5	SENTIDO	
3	TECNICA	

Figura 4 – Exemplo no modo *Menção* no WKS. Só foram rotuladas as entidades mencionadas no exemplo. Texto: NAME (2012). Fonte: Autoria própria.

Aqui cabe mencionar as peculiaridades dos textos que compuseram o conjunto de documentos. O texto acima em sua totalidade possui 771 palavras em 33 sentenças. O artista objeto da crítica é mencionado 18 vezes, e são apresentadas 5 obras específicas além de menções gerais ao seu trabalho. Tal texto exemplifica bem a estrutura geral encontrada no

¹³ Em dezembro de 2019 a IBM lançou de maneira experimental o recurso de *Relações de sentença cruzada* que permite rotular uma relação entre entidades dentro do intervalo de seis sentenças. Entretanto, tal recurso só foi disponibilizado para projetos configurados no idioma em inglês. Ver IBM Cloud Documentos, 2019b.

corpus de documentos. Sendo assim, o foco no início do treinamento do modelo foi justamente assegurar que essas entidades principais fossem treinadas de maneira satisfatória.

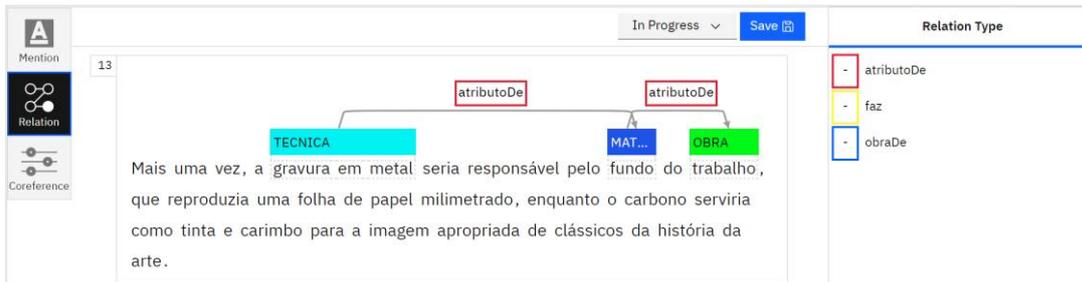


Figura 5 – Exemplo no modo *Relação* no WKS. Texto: NAME (2012). Fonte: Autoria própria.

2.5 – A ENTIDADE TÉCNICA

Por se tratar de uma pesquisa sobre arte contemporânea, a presença da *técnica* como uma das entidades à primeira vista pode parecer inadequada ou ultrapassada. E, tendo em vista, que uma das características do modelo é a sua capacidade de aprender novas informações a partir do conteúdo que lhe é dado, uma contextualização de tal escolha se faz necessária.

O primeiro ponto é a inegável herança histórica que tal termo carrega, e por mais que muitos artistas e críticos advoguem por sua rejeição ou abolição, seu uso ainda faz sentido em muitos casos. Um deles é justamente sua correspondência em alguns níveis com o termo *arte*. Nas palavras de Nunes (1989, p.17): “*Ars, artis*, a palavra latina da qual a nossa derivou, corresponde ao grego *tékne*, que significa todo e qualquer meio apto à obtenção de determinado fim, e que é o que se contém na ideia genérica de arte”. Essa relação estreita com o meio de fazer justifica o hábito de se referir a uma obra específica por sua técnica. Sendo comum que, caso seja adequado, tal obra seja mencionada como *pintura, escultura, desenho, vídeo, gravura* e assim por diante. Dada essa situação, no ambiente de anotação é possível qualificar a menção pelo contexto que esta ocorre, atribuindo assim a esta uma função de outro tipo de entidade. No caso da figura 6 a menção *pintura* apesar de rotulada como *TECNICA*, tem a função de *OBRA* dado o contexto da sentença anterior que possui outra menção desta entidade. Através de uma cadeia de correferências é estabelecido que a entidade *obra* da sentença 5 é também *pintura*, além das relações com as outras entidades na sentença 6.

O segundo ponto aprofunda a discussão mencionada anteriormente de diluição das linguagens na arte contemporânea. Conceitos como campo expandindo na obra de arte são resultados da busca por autonomia das técnicas nas vanguardas modernas. Para Elias e Vasconcelos:

O séc. XX, encarregou-se de subverter radicalmente, as hierarquias herdadas das tradições acadêmicas que respeitavam disciplinas, temas, técnicas e suportes, levando ao abandono das categorias instituídas de pintura, escultura

e desenho. As propostas artísticas interdisciplinares, onde diferentes áreas se cruzam e complementam, fertilizando-se mutuamente, tornaram obsoletas as taxonomias tradicionais. (ELIAS; VASCONCELOS, 2009, p. 3).

Apesar de uma certa torcida, ainda assim a pintura, gravura ou escultura não morreram. Portanto, pode-se dizer que o que morreu foi seu protagonismo irrevogável. Voltando ao exemplo da seção acima, a obra *Carbon copy* de Marcelo Moscheta faz um jogo com gravura em metal, transferência por papel carbono e desenho. Aqui, e em muitos casos, a interdisciplinaridade não nega as técnicas, mas justamente as incorporam. Neste sentido, é importante notar que apesar de categórica, a lógica aplicada ao modelo não é fechada ou hierárquica. E que esta, é capaz de abranger a lógica, ou as várias lógicas da arte contemporânea. Neste sentido, em seu estado ideal o modelo é capaz de identificar palavras que no texto exerçam uma função de *técnica* em relação a uma entidade *obra*, *sentido* ou *matéria* como é o caso do *fundo* de *Carbon Copy*. Ou no sentido oposto, exercendo uma função de *matéria*, como é o caso exemplificado na figura 7. Portanto, seguindo essas diretrizes, o modelo estabelecido abre espaço para uma nova taxidermia gerada a partir das próprias obras, que incorporam técnicas e materiais em estruturas nunca antes estabelecidas, e que são assentadas pelos críticos.

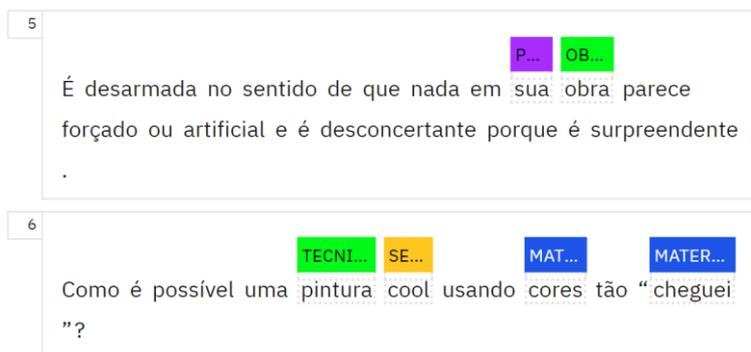


Figura 6 – Exemplo de rotulação de *função* no WKS. A cor do rotulo indica a função da menção. Texto: OSORIO (2005). Fonte: Autoria própria.



Figura 7 – Exemplo de rotulação de *função* no WKS. A cor do rotulo indica a função da menção. Texto: PALHARES (2007). Fonte: Autoria própria.

2.6 – ANOTAÇÃO, TREINAMENTO E AVALIAÇÕES

Dos 30 textos do conjunto de documentos, 21 foram designados para o conjunto de treinamento e foram pré-annotados usando três dicionários criados para os tipos de entidade

PESSOA, *TECNICA* e *OBRA*. O pré-annotador¹⁴ que utiliza o modelo padrão de *NLU* do *Watson* foi aplicado apenas ao tipo de entidade *PESSOA*, configurando que esta correspondia ao tipo de entidade padrão *PERSON*. Em seguida, foi realizada a anotação utilizando o sistema de tipos descrito no final da seção 2.3. O autor deste artigo foi o único anotador, portanto, aspectos como a concordância entre anotadores não foram considerados. O treinamento do modelo utiliza um algoritmo proprietário, e é possível baixar os conjuntos de documentos após esse processo no formato JSON já *tokenizados*. O processo de avaliação compara as rotulações realizadas no conjunto de treinamento com as aplicadas pelo modelo gerado nos conjuntos de teste e cego. Apesar do estágio inicial de desenvolvimento, e baixo número de palavras e entidades treinadas, pontuaremos as estatísticas obtidas no grupo de teste com a finalidade de registro (fig. 8), cientes que tais resultados não são necessariamente escaláveis em modelos mais complexos. Em relação aos tipos de entidade, o grupo de testes não conseguiu retornar estatísticas por conta do número baixo de densidade. Mesmo assim, o modelo conseguiu extrair as relações quando implementado, como veremos a seguir.

Tipos de Entidade	F1	Precisão	↑ Recall	% do Total de Anotações	% da Densidade do Conjunto (por n.º de palavras)
PESSOA	0.67	0.52	0.96	42% (58/139)	1% (58/4644)
OBRA	0.78	0.72	0.86	32% (44/139)	1% (44/4644)
TECNICA	0.87	0.94	0.81	27% (37/139)	1% (37/4644)
Estatísticas Gerais	0.75	0.65	0.89	100% (139/139)	3% (139/4644)

Figura 8 – Estatísticas de avaliação dos tipos de entidades no grupo de testes no WKS. Quanto mais próximo de 1.0 melhor. Fonte: Autoria própria.

3. IMPLEMENTANDO O MODELO

Para testar as capacidades do modelo, este foi implementado em um aplicativo web utilizando os recursos da API (Interface de Programação de Aplicativos) *NLU* do *IBM Watson*. O aplicativo foi desenvolvido com base no código aberto disponibilizado por Sérgio Gama (2019), e foi distribuído através da plataforma de *DevOps* (*development and operations*) da *IBM Cloud* em um domínio temporário. Na página inicial do aplicativo há um campo de entrada do texto a ser analisado (fig. 9). Abaixo dele há duas frases de exemplo. Não há limites de caracteres para o texto de entrada, sendo o formato deste UTF-8, porém o número de entidades

¹⁴ Estes pré-annotadores são opcionais e executados de maneira automática antes da anotação humana. Ver *IBM Cloud* documentos (2019a).

a serem identificadas foi limitado para 15. Também é possível selecionar se o texto será analisado utilizando o modelo customizado, desenvolvido nesta pesquisa, ou o padrão da API NLU para fins de comparação (fig.10). Ao clicar no botão de analisar o usuário é levado para uma página com os resultados tabelados (fig.11).

Análise de texto | Arte contemporânea

Ferramenta que identifica em um texto pessoas, obras, técnicas e suas respectivas relações. Utiliza um modelo de NLP customizado, treinado na plataforma Watson Knowledge Studio através de aprendizado de máquina supervisionado. Os documentos que serviram de base para desenvolvimento deste projeto são textos críticos de artistas contemporâneos brasileiros finalistas do Prêmio PIPA.

Modelo de NLP:
O modelo padrão é o do Watson NLU, o customizado é o desenvolvido neste projeto.

Customizado ▾

Insira seu texto aqui.

Texto em português, número máximo de entidades a serem identificadas limitado para 15.

Analisar

Textos de exemplo

A exposição "Fabricio Lopez – Várzea" com trabalhos inéditos do artista paulista que oferece novas dimensões à prática de gravura em madeira. [fonte](#)

Figura 9 – Página inicial do aplicativo web desenvolvido. Fonte: Autoria própria.

Texto analisado

"A exposição "Fabricio Lopez – Várzea" com trabalhos inéditos do artista paulista que oferece novas dimensões à prática de gravura em madeira."

Entidades

Tipo	Texto	Confiança	Contagem
Person	Fabricio Lopez	0.978348	1

Relações

1ª Entidade	Relação	2ª Entidade	Relevância
Person: Fabricio Lopez	locatedAt	Location: Várzea	0.24353
Person: artista	employedBy	Location: paulista	0.264623

Figura 11 – Resultado do texto de exemplo utilizando o modelo padrão. Fonte: Autoria própria.

Entidades

Tipo	Texto	Confiança	Contagem
PESSOA	Fabricio Lopez	0.995151	2
TECNICA	gravura	0.973926	1
OBRA	trabalhos	0.676838	1

Relações

1ª Entidade	Relação	2ª Entidade	Relevância
OBRA: trabalhos	obraDe	PESSOA: Fabricio Lopez	0.693275

Figura 12– Resultado do texto de exemplo utilizando o modelo customizado. Fonte: Autoria própria.

Uma questão que se sobressaiu ao utilizar o aplicativo foi a quantidade do texto de entrada. Observamos que uma frase é mais bem analisada quando há outras sentenças com as mesmas entidades no texto de entrada, se comparado com essa mesma frase sendo analisada individualmente. Porém, em um texto completo com uma média de 1000 palavras quando

analisado o resultado apresentava com frequência *artista* e nomes próprios como entidades diferentes, apesar das entidades pontuarem uma contagem alta.

4. CONCLUSÕES E POSSIBILIDADES FUTURAS

Este artigo apresentou o início do desenvolvimento de um modelo de compreensão de linguagem natural treinado a partir de aprendizado de máquina, e sua aplicação em uma ferramenta com funções básicas de análise de textos sobre arte contemporânea. Em desdobramentos futuros, foram pensadas implementações de um modelo como este em aplicações que fazem uso das funções de identificação de entidades mencionadas e extração de relações de maneira mais específica. No campo da pesquisa, em arte há a opção de uma ferramenta integrada a domínios e acervos digitalizados. Os documentos e textos ali poderiam ser estruturados a partir das entidades identificadas. Em adição a catalogação tradicional, filtros complexos poderiam ser aplicados cruzando artistas, linguagens e temas de interesse. A integração do modelo desenvolvido a um serviço como o *IBM Watson Discovery*, possibilitaria a criação de uma aplicação capaz de responder perguntas específicas do campo da arte. Tais perguntas poderiam ser realizadas tanto por meio de voz ou texto por um público mais leigo no contexto da mediação de exposições de arte. Já no campo da produção textual, aliado a aplicações de sumarização, um texto poderia ser resumindo mantendo aspectos mais relevantes definidos pelo usuário, por exemplo por artista ou técnica específica. Outra possibilidade seria compilar vários textos de um mesmo artista, gerando um novo texto que conte sua trajetória a partir de uma organização de suas obras.

Por fim, o presente artigo contribui para aproximação dos campos de pesquisa sobre arte, computação e inteligência artificial. Campos estes que ganharam novos sentidos e relações devido ao contexto de pandemia e isolamento social que se estabeleceu durante o desenvolvimento desta pesquisa. Tendo em vista a tendência das mediações que uma obra de arte pode ter na internet e sua presença nos meios digitais, e reforçando o aspecto da arte contemporânea de contato com a esfera cotidiana. Sendo assim, para trabalhos futuros tal contexto deverá ser considerado no emprego de tais tecnologias na criação de ferramentas que visem aproximação da arte contemporânea com a esfera social.

REFERÊNCIAS

AMADO, Guy. *Notas sobre a jovem crítica de arte*. In: FERREIRA, Glória. (Org.). *Crítica de Arte no Brasil: Temáticas Contemporâneas*. Rio de Janeiro: Funarte, 2006. p. 243 – 245.

COCCHIARALE, Fernando. *Crítica: a palavra em crise*. In: FERREIRA, G Glória. (Org.). *Crítica de Arte no Brasil: Temáticas Contemporâneas*. Rio de Janeiro: Funarte, 2006. p. 217 – 220.

ELIAS, Helena; VASCONCELOS, Maria. *Desmaterialização e campo expandido: dois conceitos para o desenho contemporâneo*. 2009.

GAMA, Sérgio. *wks-tooling*. GitHub, 2019. Disponível em: <https://github.com/sergiogama/wks-tooling>. Acesso em: 20 de ago. 2020.

IBM Cloud documentos. *Fluxo de trabalho de criação de modelo de aprendizado de máquina*. 2019a. Disponível em: https://cloud.ibm.com/docs/services/watson-knowledge-studio?topic=watson-knowledge-studio-ml_annotator&locale=pt-BR. Acesso em: 05 de mai. 2020.

IBM Cloud documentos. *Notas sobre a liberação*. 2019b. Disponível em: <https://cloud.ibm.com/docs/watson-knowledge-studio?topic=watson-knowledge-studio-release-notes&locale=pt-BR>. Acesso em: 08 de ago. 2020.

IBM Cloud documentos. *Incluindo documentos para anotação*. 2019c. Disponível em: <https://cloud.ibm.com/docs/services/watson-knowledge-studio?topic=watson-knowledge-studio-documents-for-annotation&locale=pt-BR>. Acesso em: 04 de mai. 2020.

Instituto PIPA. Prêmio PIPA 2019. Rio de Janeiro: Instituto PIPA, 2019. 272 p. (catálogo).

JURAFSKY, Daniel; MARTIN, James. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. 2ª. ed. Upper Saddle River: Prentice-Hall, 2008.

NAME, Daniela. *Marcelo Moscheta E O Pico Do Jaraguá*. Texto curatorial da exposição NORTE, realizada no Paço Imperial, no Rio de Janeiro, 2012. Disponível em: <https://www.marcelomoscheta.art.br/marcelo-moscheta-e-o-pico-do-jaragua>. Acesso em: 6 de ago. 2020.

NUNES, Benedito. *Introdução à Filosofia da Arte*. São Paulo: Ática, 1989.

OSORIO, Luiz Camillo. *Por um belo desconcertante*. Ensaio escrito para exposição Galeria Virgilio, São Paulo. 2005. Disponível em: <http://www.tatianablass.com.br/textos/205>>. Acesso em: 6 de ago. 2020.

PALHARES, Taísa. *Zona Morta*. Escrito para a exposição no Centro Universitário Maria Antonia-USP, São Paulo. 2007. Disponível em: <http://www.tatianablass.com.br/textos/203>. Acesso em: 6 de ago. 2020.