

Interfaces Contemporâneas no Ecossistema Midiático

Rodrigo Cunha e Taciana Burgos (orgs)

Prefácio: Carlos Scolari



Interfaces Contemporâneas no Ecosistema Midiático. Taciana Burgos & Rodrigo Cunha.
- 1ª Edição - Aveiro: Ria Editorial, 2019.
169 p.

Livro digital, PDF.

Arquivo Digital: download e online
Modo de acesso: www.riaeditorial.com
ISBN 978-989-8971-02-9

Prefácio de Carlos A. Scolari

1. Interface. 2. Ecosistema Midiático. 3. Narrativa Transmídia. 4. Convergência
Midiática. 5. Regimes de Interação. I. Burgos, Taciana. II. Cunha, Rodrigo. III. Título.

Ficha catalográfica: 19/02/2019

Copyright das imagens pertencem aos seus respectivos autores.

© Design de Capa: Denis Renó
Paginação: Luciana Renó

© Taciana Burgos e Rodrigo Cunha (Organizadores)

© Prefácio de Carlos A. Scolari

© Alexandra Fante, Alfredo Caminos, André Fagundes Pase, Carlos Eduardo Marquioni, Carolina Weber Dall'Agnese, Claudia Ardini, Edson Capoano, Eduardo Campos Pellanda, Esau Salvador Bravo Luis, Eugenia Mariano da Rocha Barichello, João Canavilhas, Jonas Gonçalves, Mágda Rodrigues da Cunha, Maria Clara Aquino Bittencourt, Melissa Streck, Nahum Mirad, Rodrigo Cunha, Taciana Burgos, Tássia Becker Alexandre, Yasmín Sayán Casquino



Licença:

>: Atribuição-Não Comercial-Sem Obras Derivadas 4.0 Internacional
>: Você é livre para:
- copiar, distribuir, exibir, e executar a obra

Baixo as seguintes condições:

- Atribuição. Você deve atribuir a obra na forma especificada pelo autor ou o licenciante.
- Não Comercial. Você não pode usar esta obra com fins comerciais.
- Sem Obras Derivadas. Você não pode alterar, transformar ou criar sobre esta obra.

<https://creativecommons.org/licenses/?lang=pt>

ISBN 978-989-8971-02-9

<http://www.riaeditorial.com>

A Voz como Interface no Sistema Ecoevolutivo dos Assistentes Pessoais

André Fagundes Pase
Eduardo Campos Pellanda
Mágda Rodrigues da Cunha
Melissa Streck

As transformações tecnológicas geram desconfiança até o momento em que os sujeitos compreendam e vejam vantagens no uso de suas propriedades. No processo desconfiança-apropriação evidenciam-se debates para além dos propostos originalmente. O trajeto é dinâmico e ocorre no jogo de tentativas, erros, acertos e adaptações tanto da indústria quanto do público. Inseridos em um mesmo ecossistema, constroem soluções no diálogo projeto e apropriação. Neste texto, o objetivo é analisar o tema a partir de investigação envolvendo os assistentes pessoais¹, nos quais há uma interface ubíqua, com comandos controlados por voz. Tal interface está em evolução e se amplia para a ubiquidade doméstica. Serão descritos o momento evolutivo destas tecnologias, ainda em fase de desenvolvimento e desconfiança, simultaneamente, seu funcionamento e os dilemas, debates e tendências deste diálogo constante entre designers e usuários, como pensa Scolari (2018). A pesquisa apresentada neste texto foi realizada pelo laboratório Ubilab² junto a uma empresa de conteúdo digital e resultou no desenho de solução para assistentes virtuais. No relato das etapas do experimento, é possível compreender a necessidade de novos caminhos metodológicos e abordagens interdisciplinares que os formatos emergentes suscitam.

A Histórica Construção da Interface

O desenvolvimento de soluções complexas e a construção de interfaces entre humanos e máquinas, que chegam ao tempo presente, traçam uma trajetória de tentativas, erros e acertos, programação, desconfiança, aceitação e novas transformações. No diálogo permanente com o usuário, a indústria tenta, cada vez mais, prever como serão as conversações que usam a voz. O momento pode parecer o ápice desta tentativa, uma vez que prever um diálogo por voz trate se processo altamente complexo.

1 O termo “assistente pessoal” pode ser designado tanto no gênero feminino quanto no masculino. Para este trabalho, mantivemos o gênero masculino por ser comumente utilizado pela empresa para a qual foi realizada a pesquisa junto ao Ubilab.

2 O Ubilab é um laboratório de pesquisas aplicadas do Programa de Pós-Graduação em Comunicação da PUCRS. www.pucrs.br/famecos/ubilab

Até chegar a este ponto, porém, há um caminho a ser narrado, destacando as investidas da indústria.

O ano de 2007 marcou não só o lançamento do iPhone mas também o início de uma transformação profunda na forma como percebemos a internet. Passados mais de dez anos, vemos uma década densa em aceleração tecnológica pautando novos mercados econômicos e comportamentos sociais. A rede ubíqua e *always-on* (Pellanda, 2005) trouxe as interações sociais virtuais para onde elas acontecem: as ruas das metrópoles, como previa Mitchell (2003). A convergência destas relações virtuais e atuais (Lévy, 1999) transformaram também as cidades e os espaços pela potência de conexão, como também descreveu Mitchell (2003). Essa relação foi amplificada pelos softwares em forma de Apps, que passam a resolver problemas específicos do cotidiano (Manovich, 2013). Esse cenário faz com que o *smartphone* se torne o principal produto mercadológico do século XXI, ao atingir a marca de 1.4 bilhões de unidades vendidas ao ano em 2018³.

Antes do iPhone, muitos fabricantes, como Nokia, Pam ou Motorola propõem o *smartphone* em diversas formas, com teclados físicos em formatos variados, bem como telas em proporções diversas. A Apple trouxe todos os elementos físicos essenciais para o *software* dentro da tela e desenhou o aparelho em torno disso. O design monolítico se tornou icônico e pautou o resto da indústria com a expansão do sistema Android da Google. Mas o objeto *smartphone*, como proposto, não consegue preencher todas as oportunidades de interação, deixando vácuos parcialmente preenchidos ao longo dos últimos dez anos. O iPad, em 2010, foi a expansão do smartphone em direção ao *laptop* e ocupa o espaço de consumo de mídias, provocando novas formas de distribuição de conteúdos para meios como jornais e revistas. Os *wearables*, ou computação vestível, começam a explorar novas fronteiras de comunicação direta com elementos do corpo. O *Google Glass* foi um experimento que virou produto para mercados verticais, explorou interações com a retina, através de sensores de percepção da íris e câmera fotográfica com ângulo de visão na altura dos olhos. Este dispositivo buscava novos usos em relação ao *smartphone* bem como a eliminação da necessidade de conferência repetitiva de mensagens nas telas dos telefones. Tais características foram transportadas pela Google para os *smartwatches* e adotadas pela Apple como forma de desenvolver dispositivos que se comuniquem com aspectos vitais do corpo.

Paralelo a isso, o barateamento de componentes que conectam objetos à internet inaugura uma nova categoria: a Internet das Coisas (*Internet of Things ou IoT*), com potencialidades de gerar outras interações de comunicação tanto de “coisas com coisas” como “coisas com pessoas”. Estudos no campo da tecnologia, como o de Hammoudeh e Ariuna, (2018) buscam demonstrar que o mundo está entrando em uma nova era de inovação, contando com diferentes sensores sem fio, além de processadores menores, mais potentes e dispositivos móveis mais inteligentes e de baixo custo. São mencionados ainda o armazenamento em nuvem, novos níveis de conectividade pela internet e *big data analytics* como potencializadores de tecnologias M2M (Máquina para Máquina). Portanto, mesmo produtos já consolidados, como os *smartphones*, acabam sendo alvo de novas demandas de consumo nesta expansão tecnológica.

Pensando em ampliar as formas de interação do *smartphone*, a Apple introduziu em 2011 o

3 Dados publicados pela empresa de análise de mercado Statista, recuperado de <https://www.statista.com/statistics/263441/global-smartphone-shipments-forecast/>

assistente pessoal Siri no iPhone 4s. Esta tecnologia vem da aquisição, em 2010, da *startup* também chamada Siri, que desde 2007 trabalhava em sistemas de reconhecimento e sintetização de voz, com fundos da agência de defesa norte-americana (DARPA). A Apple basicamente começou a desenvolver lentamente o seu sistema de assistente pessoal para que abrangesse mais línguas, antes de novas funcionalidades.

Em outro sentido, a Amazon, segunda empresa a obter o valor de mercado de 1 trilhão de dólares depois da Apple, vinha buscando espaço através de algum dispositivo pessoal que lhe oferecesse mais relevância já que marcava liderança junto aos leitores de livros digitais, com o Kindle. No final de 2014 lançou o *Amazon Echo*, um alto-falante em 360 graus, com diversos microfones espalhados em volta do cilindro, para viabilizar um assistente pessoal chamado Alexa. Este produto resultou de pesquisa do laboratório da empresa denominado Lab126, localizado no Vale do Silício e em Cambridge, Massachusetts. A Amazon já tinha tentado, sem sucesso, entrar no segmento dos *smartphones* e dos *tablets* e viu neste produto a oportunidade de abrir um novo mercado.

Diferente da Siri, o *Amazon Echo* esteve disponível para interações nas casas com o comando através da palavra-chave Alexa. A eficiência do aparelho e do sistema fez ele se expandir rapidamente e resultou em uma linha de produtos da Amazon. A Google também viu a oportunidade de apresentar o seu produto de assistente pessoal em forma de alto-falante, assim como a Apple, em um segundo momento, com o Home Pod.

O mercado de assistentes pessoais ganha um novo sentido fora dos *smartphones* e os assistentes pessoais começam a ser os responsáveis pela interação do homem com o computador, através da voz. Algo que estava no imaginário coletivo provavelmente desde o filme 2001: Uma Odisseia no Espaço, de Stanley Kubrick, de 1968. Ao mesmo tempo em que estes aparelhos simplificam e personalizam a interação homem/máquina, representam sistemas complexos de computação em nuvem que processam um número significativo de operações por segundo, para que a comunicação exista. O ponto chave passa a ser o diálogo em forma de interface homem/máquina. Fatores derivados do histórico da conversação e informações de contexto do indivíduo precisam ser processados em tempo real. A interação, em parte, deve ser desenhada e prevista por programadores e, em parte, processada por algoritmos. Nesta etapa, faz-se necessário o desenvolvimento de tecnologias de Inteligência Artificial como o *Deep Learning*, que processa contextos diversos para encontrar padrões que combinem com a questão solicitada.

Para os produtores de conteúdo deste novo ambiente cabe, agora, a tarefa de desenhar não necessariamente o final da interação, mas cada vez mais entender o meio. Como podem ocorrer os caminhos de conversações ou como serão os contextos em que o indivíduo vai fazer uma requisição são perguntas-chave para o entendimento do processo. Por precisar antever a pergunta, processá-la com contextos e entregar uma resposta pertinente, o desenvolvimento da arquitetura destes sistemas deve envolver equipes cada vez mais multidisciplinares. A simplicidade aparente da interação sem a interface visual encobre uma complexa rede de processamentos, resultado do estado da arte da ciência da computação, bem como da evolução dos interagentes que operam estas comunicações. Ao “conversar” com a notícia ou com a informação, o indivíduo precisa ter vivenciado diversas fases de

literacia midiática⁴ que o capacitaram para este momento.

O Diálogo como Lei na Interface

Nesta literacia, entretanto, evidencia-se o jogo da pergunta e da resposta que, como citado anteriormente, leva a indústria a tentativas, erros e acertos. Não é possível imaginar que o usuário responderá sempre da forma como a interface foi planejada. Todos atuam e interagem, neste amplo ecossistema, do qual a indústria é parte importante, mas onde o usuário desenha táticas em resposta às estratégias apresentadas, o que torna a interface um ambiente-chave para qualquer investigação que envolva todos os atores.

É sob a influência de Pierre Lévy, Kevin Kelly, John Maeda e Marshall McLuhan que Scolari (2018) vai propor dez leis da interface. Neste texto, o objetivo é abordar a evolução da interface ubíqua, através da qual diferentes sistemas são controlados pela voz. É um processo que se amplia para a ubiquidade doméstica e no qual a construção de Scolari (2018) se faz adequada, por defender que as interfaces se desenvolvem no diálogo entre a interface e o usuário.

A inspiração em McLuhan vem da obra produzida pelo autor canadense com o filho Eric, intitulada *As leis da mídia*, com o objetivo de reduzir a complexidade da esfera da mídia a um conjunto de princípios resumidos por Scolari (2018): a extensão, aquilo que o meio amplifica ou intensifica, a obsolescência, quando se torna arcaico, a recuperação, quando recupera o que havia perdido anteriormente e a reversão, o que acontece com o meio quando levado aos seus limites. Em Maeda, busca a simplicidade, que “consiste em subtrair o que é óbvio e adicionar o específico”. Scolari (2018) admite que a obra vai muito além dos autores citados e transita entre outros nomes, disciplinas e teorias. E é exatamente por esta característica de transitar e conectar teorias, para solucionar questões que envolvem a transformação do ecossistema da mídia, que Scolari se faz pertinente na análise do tema dos assistentes pessoais, cuja investigação determina uma visão teórico-metodológica interdisciplinar.

O autor propõe uma aproximação do biológico com o tecnológico, numa aplicação da esfera bionatural na esfera tecnocultural. Reconhece que, no modelo ecoevolutivo, as transformações não são diretas e cita como etapas o período de valorização da estrutura (1950), do signo (1960), do discurso (1980) e sugere que a Interface seja o conceito fundamental do século XXI. Mas por que leis? É o entendimento como um conjunto de relações entre elementos que expressam uma regra ou comportamento de um fenômeno. Na ação de desconstrução e construção de conceitos, o autor apresenta suas 10 leis e afirma que a interface é um lugar de interação. As interfaces não são transparentes, conformam um ecossistema, evoluem, coevoluem com seus usuários, não se extinguem, mas se transformam, se não podem fazer algo, simulam, estão submetidas às leis da complexidade. Defende ainda que o desenho e o uso de uma interface são práticas políticas e que a interface é um lugar de inovação.

Scolari (2018, p. 226) enfatiza as conversações constantes entre desenhista e usuário, numa perspectiva semiótica na qual o leitor dá sentido ao texto. “A interface é a mediadora de um intercâmbio

4 Capacidade de gerir informações e analisar mensagens recebidas através de mídias diversas, além de criar conteúdos de qualidade e distribuí-los de forma coerente com o contexto digital, contribuindo com o diálogo social.

bio que funciona de maneira muito similar a relação entre autor-texto-leitor.” Considera a interface um lugar de interação e indica que a melhor é aquela que facilita a conversação, a que desaparece e permite foco na atividade central do usuário. Pensa que em praticamente todas as áreas muitos são desenhistas, pois criam ambientes de interação. Dando continuidade a sua sugestão de que este seja um conceito central para o século XXI, o estudioso argentino trabalha, desde o início da obra, com metáforas para construção de suas dez leis. A interface, porém, é a “mãe” de todas as metáforas, a mais popular, a primeira que um usuário vai elaborar frente a um artefato tecnológico. A interface é um espaço de interação e sua criação só ocorre quando vários atores entram em contato e interagem entre si. Mas o que existe é uma coevolução das interfaces com seus usuários.

A coevolução conecta dois ou mais elementos de um mesmo domínio ou domínios diferentes. Domínios são conjuntos coerentes de artefactos, métodos e práticas que não se criam, emergem. Cada domínio instala um vocabulário e uma linguagem baseados em suas próprias regras, como foram a imprensa ou os aviões (Scolari, 2018). O autor escreve também sobre inovação radical, quando uma mudança é feita a partir de uma ruptura, e inovação incremental, modificações que ocorrem a partir de pequenas alterações em objetos já existentes. A web colaborativa é uma nova configuração, pensa Scolari (2018), das relações de atores humanos e tecnológicos, que gera uma nova interface. No caminho da inovação, é fundamental ter sempre em conta as táticas dos usuários, já que todas as estratégias propostas são negociadas e reinterpretadas pelo usuário dentro da interface.

Reconhecendo os riscos do cruzamento das esferas biológica e tecnológica ou mesmo que pode ser problemático trabalhar com analogias, o autor sugere cautela devido às grandes diferenças entre o domínio biológico e o tecnológico. Scolari (2018) propõe o conceito de coevolução, num processo dinâmico de desenvolvendo em que todos conversam e fazem evoluir as interfaces. Tradicionalmente, pensa ele, as investigações estão centradas nos artefatos da evolução tecnológica, enquanto outras análises trazem os inventores como protagonistas de seu modelo analítico ou mesmo as forças sociais. Em seu livro, a sugestão é que a interface seja a unidade de análise. A postulação é considerar um ecossistema, inserindo artefatos, inventores e forças sociais em uma rede sociotécnica de relações, intercâmbios e transformações, para analisá-los de uma perspectiva coevolutiva.

A Voz como Interface

Seguindo o pensamento de Scolari (2018), nos vemos em um ecossistema no qual convivemos com tecnologias que, há meio século, eram componentes de cenários de filmes de ficção científica e estavam sendo premeditadas pelos estudos de McLuhan (1964). O filme 2001: Uma Odisseia no Espaço (1968) mostra parte da evolução humana e o surgimento de instrumentos de comunicação entre os seres humanos pré-históricos, devido a sua nova capacidade de imaginar. É através de grunhidos que um bebê, junto a gestos, começa a comunicação para expressar suas necessidades. Com o passar dos anos, a fala evolui e surgem os idiomas para diferentes culturas. A língua, no sentido do idioma, preenche lacunas de comunicação entre indivíduos e até culturas. O filme ainda sugere que a forma mais imersiva de falarmos com as máquinas seria com diálogos pela voz. Ao dar também um nome ao computador, HAL, Arthur C. Clarke e Stanley Kubrick, criadores do filme, personificam o diálogo.

Possivelmente, por este motivo, Siri (Apple), Cortana (Microsoft) e Alexa (Amazon) tenham nomes de pessoas. A especulação da derivação direta do filme pode não ser completamente verossímil. De qualquer forma, demonstra o imaginário coletivo sobre o tema.

A personificação dos assistentes pessoais por voz também pode ser entendida como metáfora. Rogers, Sharp e Preece definem que “as metáforas de interface têm a intenção de fornecer entidades familiares que permitam às pessoas facilmente compreenderem o modelo conceitual subjacente e saberem o que fazer em uma interface” (Rogers, Sharp & Preece, 2011, p. 44). Conforme Scolari (2018, p. 30), uma nova metáfora pode surgir a qualquer momento, iluminando outros aspectos, privilegiando uma visão sobre algumas propriedades e ocultando outras características da interface⁵. Ou seja, as interfaces, mesmo por voz, podem ser moldadas e redefinidas, ganhando novos sentidos e novos tipos de uso.

Quando elucida a palavra falada, em capítulo específico, McLuhan aborda a questão da linguagem como protetora e ampliadora do homem, mas comenta que ela também divide suas faculdades. “A consciência coletiva e o conhecimento intuitivo ficam diminuídos por esta extensão técnica da consciência que é a fala” (McLuhan, 1964, p. 97), ou seja, a fala é sobreposta aos pensamentos humanos. Sem a fala, o homem estaria preso somente a suas próprias ideias, não permitindo gerar um desenvolvimento comum em sua sociedade. O autor compara a importância da fala, em relação ao pensamento, com a invenção da roda em relação aos pés, e mais adiante comenta que, através da fala, é possível a existência de diferentes idiomas e estilos, como os existentes na moda e na arte. Cria-se uma sociedade organizada em grupos identificados por algum tipo de linguagem.

Ao analisar a oralidade e sua presença na relação com a escrita, Ong (1998) pensa que a fala é inseparável da consciência e tem fascinado os seres humanos desde os mais antigos estágios desta mesma consciência. A mesma fascinação pelo discurso oral segue inalterada séculos depois de a escrita ter sido posta em uso (Ong, 1998). O autor afirma que ler um texto significa convertê-lo em som, em voz alta ou na imaginação e entende que a escrita “tiranicamente” encerra as palavras num campo visual. Na constante defesa da importância da manifestação oral, o autor indica que os seres humanos se comunicam de muitas maneiras, fazendo uso de todos os seus sentidos. Contudo, entende que, num sentido profundo, a linguagem, o som articulado, “tem importância capital” (Ong, 1998, p. 15). Não apenas a comunicação, mas o próprio pensamento está relacionado de forma especial ao som e, ao citar Siertsema (1955), defende que, onde quer que existam seres humanos, eles têm uma linguagem que se constitui basicamente por ser falada e ouvida, no mundo sonoro.

Ong (1998) reconhece os valores da cultura escrita, sem os quais a consciência nunca alcançaria estágios altamente interiorizados nos quais o indivíduo está imerso inconscientemente. A interação, pensa ele, entre a oralidade na qual todos os seres humanos nascem e a tecnologia da escrita, na qual ninguém nasce, atinge as profundezas da psique. “Onto e filogeneticamente, é a palavra falada que primeiramente ilumina a consciência com a linguagem articulada, a primeira que divide o sujeito e o predicado e depois os relaciona entre si, e que estabelece laços entre os seres humanos na sociedade.” (Ong, 1998, p. 199)

5 Tradução própria para o original: “En cualquier momento puede aparecer una nueva metáfora que ilumine otros aspectos, privilegiando una mirada sobre ciertas propiedades y ocultando otras características de la interfaz.”

Mas se Ong (1998) vai lembrar que de todas as dezenas de milhares de línguas faladas no curso da história humana, somente cerca de 106 estiveram submetidas à escrita num grau suficiente para produzir literatura e a maioria jamais foi escrita, McLuhan busca considerar a codificação no ecossistema da computação. Aponta que, com o surgimento dos computadores, as línguas passam a se codificar, podendo ser traduzidas através de códigos para outras línguas. Isto teria criado uma nova condição de compreensão e unidade universal (McLuhan, 1964, pp. 68-69). Porém, tendo os computadores se transformado em elementos cada vez mais pessoais, atualmente na forma de *smartphones*, que são levados a todos locais nos bolsos das pessoas, esta codificação está cada vez mais fragmentada. Surgiram apps que atendem as mais específicas necessidades, e a interface por voz já é uma realidade para executar diferentes ações no contexto digital.

Em uma perspectiva mais técnica, Moorty e Vu (2015, p. 307) apontam que estes recursos de interface têm se aprimorado, tornando possível a inserção de comandos através de voz, e a invenção de assistentes pessoais ativados pela fala, definindo uma característica padrão para muitos dispositivos móveis. Os autores ressaltam que as principais plataformas de telefonia móvel já possuem algum recurso nativo de assistente por voz (VAPA) em seus *smartphones*: Apple - Siri; Google - Google Now; Samsung - Samsung Voice. Em termos mais técnicos, os assistentes deste tipo, cuja interface entre dispositivo e usuário é conhecida pelo termo *Voice User Interface* (sigla VUI), e permitem diálogos entre humanos e máquinas, conforme Lotterbach & Peissner (2006), levam em consideração regras pré-definidas. Segundo os autores, há restrições que evidenciam as limitações da tecnologia em relação a questões da fala. Neste sentido, sotaques e variações de palavras podem não ser entendidas pela máquina, gerando um tipo de quebra no discurso.

A Voz como Interface na Pesquisa Aplicada

O processo deste desenvolvimento é dinâmico, conforme apresentado anteriormente. Muitas são as camadas de investigação e tentativas até chegar a uma proposição com sentido aos sujeitos que farão sua reinterpretação na interface, como propõe Scolari (2018). Neste capítulo, utilizamos como uma destas camadas a pesquisa utilizando interface de voz para o G1⁶, canal de notícias da empresa brasileira de conteúdo digital Globo.com⁷, realizada no segundo semestre de 2017, pelo Ubilab. A partir de uma provocação inicial, o projeto, com duração de seis meses, diagnosticou possibilidades para o uso desta tecnologia na comunicação.

O desenvolvimento do protótipo foi realizado em sete etapas: estudo inicial, estudo de dinâmicas de linguagem, análise das rotinas de produção, desenvolvimento de protótipo, testes iniciais, visita técnica ao G1 e entrega do relatório. Para tanto, uma equipe de sete pessoas foi dividida em dois times, um com foco na programação e outro dedicado a compreender como as notícias poderiam ser evocadas (para utilizar a expressão correta neste contexto) e estruturadas para que a Inteligência Artificial do Google processasse e informasse o público. Os três passos finais não serão detalhados neste

6 Site noticioso brasileiro, disponível online em <https://g1.globo.com/>.

7 <https://www.globo.com/>

capítulo, por compreenderem conhecimento específico da empresa e, sobretudo, por não agregarem informações relevantes ao estudo.

A primeira etapa foi a análise do cenário para esta comunicação baseada em voz. Conforme observado no desenvolvimento destas soluções, muitas são as tentativas. Algumas permanecem e outras são apenas estratégias de marketing que também não seguem adiante. Entre outras iniciativas, no momento era possível dividir em três possíveis caminhos para o projeto vinculado com interfaces específicas, Siri (Apple), Alexa (Amazon) e Home (Google). Em virtude da disseminação do sistema Android e da possibilidade de utilizar os assistentes para *smartphones* em paralelo com o aparelho Google Home, este foi o caminho escolhido. Outro fator relevante era a parceria já existente entre a Globo.com e a empresa norte-americana, em virtude do uso prévio de diversas outras tecnologias.

As notícias tradicionais foram definidas como assunto prioritário. Esta editoria é envolvida por rotinas de produção de conteúdo atualizado em tempo real, além do acesso de uma forma diferente. Seria possível utilizar a interface para descobrir informações sobre eventuais capítulos de novela, por exemplo, porém dados pontuais como pesquisas eleitorais e últimas notícias permitem compreender o uso da assistente como um mordomo disposto a oferecer a informação. Isto é extremamente útil em um momento de acesso tão rápido, que não pode passar por uma timeline ou visita a um website.

Também foram observados outros exemplos de transposição de meios não-sonoros para o Google Home. É possível destacar dois usos simbólicos. O primeiro pode ser apontado como estratégia, enquanto o outro foi aplicado como recurso apenas para divulgação. Ambos envolvem o meio jornalístico, sendo corporações que buscam novos espaços no ecossistema midiático em função das novas tecnologias. Trata-se do jornal New York Times e da revista Vogue, ambos americanos. A seguir a experiência de cada um com os assistentes pessoais.

Além de ser referência mundial como jornal, o New York Times também é conhecido por experimentar novas possibilidades de comunicação. Cabe recordar que foi utilizado como exemplo no lançamento do iPad e também distribuiu *cardboards* para os assinantes utilizarem seu aplicativo de vídeos em 360°. No caso do Google Home, a estratégia adotada foi utilizar o aparelho para ouvir os podcasts produzidos pelo jornal. Além disso, no final de 2017, ao assinarem o jornal por um ano, as pessoas poderiam adquirir um assistente, com desconto.

Neste caso, o assistente é posicionado como uma forma de automatizar rotinas domésticas a partir das vantagens para quem é assinante. O New York Times não apenas coloca em prática um uso de vanguarda da informação digital, mas também auxilia na formação do ecossistema do seu uso (Garun, 2017). Esta utilização lembra a realizada por empresas como a BBC na época da implantação da TV Digital no Reino Unido, quando a emissora também informava e estimulava adesão a nova tecnologia. A estratégia, portanto, compreende em oferecer vantagens no conteúdo e no custo para ter o assistente.

A edição especial de 125 anos da revista Vogue (Rugel, 2017; Vogue, 2017) contou com uma aplicação específica para o uso do Google Home. O leitor poderia ouvir comentários da equipe do magazine sobre cinco reportagens, como em uma espécie de extensão do conteúdo. Estes cinco áudios duravam poucos minutos.

Embora as rotinas de comando da aplicação fossem práticas e simples, o material só foi utilizado desta vez. A ideia de complemento foi interessante, porém seu uso pouco estimulado. Possivelmente será lembrado não como uma forma da revista dialogar com seus leitores, mas apenas um caso de

marketing através da nova ferramenta, desperdiçando boa possibilidade de estimular um novo canal com seu público. O Google conta ainda com mais de 100 aplicações de uso com a Home. A maior parte delas é para ouvir podcasts ou boletins com horário de publicação definida com regularidade.

Compreendido o contexto, a pesquisa dividiu esforços. Semanalmente, porém eram realizados encontros entre os grupos de programação e desenvolvimento de estruturas de linguagem. Desta forma, foi fomentado o compartilhamento de descobertas e um grupo auxiliava o outro com percepções sobre o conhecimento. Paralelo a este trabalho, outro grupo buscou compreender como funcionam os processos baseados nas chamadas vocais que envolvem a compreensão das notícias. Por ser realizado entre homens e máquinas, isto abrange diferentes concepções do processo de diálogo.

Este não é o caso de produzir uma notícia para que outra pessoa leia ou um boletim redigido para que uma interface de voz transforme um texto em som. Há um processamento de instruções de entrada pelo homem, análise pela máquina e consequente construção de um bloco de informações que será lido pela interface de voz e deve ser compreendido por um agente humano.

Assim, foi necessário entender a importância da linguagem, perspectiva histórica oferecida por Pinker (2004), além de um estudo das bases da linguística (Saussure, 1987). Em virtude da concepção nova da plataforma, este conhecimento foi transposto para a informática e suas estruturas de processamento e, sobretudo, desenho de interface. Desta forma, Laurel (1990) e Moggridge (2007) contribuem para este processo, sobretudo na ênfase da pesquisa por uma interface simples e eficiente, diante das dificuldades impostas pela língua falada.

O trajeto da notícia desenvolvido para este projeto parte de uma ordem que aciona o Google, posteriormente outra para um serviço específico e aí sim, de fato, o processo da notícia. As dificuldades aumentam exponencialmente a cada nova camada. Enquanto a primeira envolve acionar o serviço, o padrão “ok, Google”, depois é preciso utilizar uma palavra-chave ou expressões que indiquem quais assuntos ou fontes de informação serão utilizados. É preciso que o sistema entenda que a expressão “dê as novas”, por exemplo, não seja algo genérico, mas sim combinado com uma fonte produtora de material de um assunto específico e em sintonia com as preferências do usuário. Este passo não forma apenas uma função de mera adaptação para as preferências, mas aproxima a máquina como uma ferramenta útil na experiência do humano. O mordomo robô não apenas entrega, mas oferece a informação ideal da maneira mais rápida possível.

Dois pontos podem ser definidos como complicadores nesta espécie de equação linguística. O primeiro é a possibilidade de utilizar preferências da plataforma específica. O Google, porém, não permite algumas combinações de serviços como preferências por endereços ou a agenda. Assim, todas as informações sobre uma equipe de futebol ou um político em específico devem ser processadas pela fonte de conteúdo.

A outra é a diversidade de sotaques e expressões utilizadas para descobrir fatos novos e acionar de fato este serviço de informações. A pesquisa inicial mapeou cerca de 40 expressões utilizadas, número que aumenta com a possibilidade de incorporação de gírias regionais para um eventual serviço. Esta riqueza de sotaques é uma dificuldade. Solucionar isto no contexto brasileiro, entretanto, permite desenvolver um projeto mais robusto. Assim, a proposta buscou estrutura vocal que seja instrução tanto humana, pois quem ordena deve entender o que está solicitando, mas que seja decodificada pela máquina, em uma espécie de aprendizado constante e duplo. A ideia pode ter uso em projetos noticio-

tos e eventuais adaptações. Para tanto, foi utilizada a estrutura ENGATE + VERBO + NOTÍCIAS + CONTEXTO/ESPECÍFICO, que permite acionar a ferramenta e granular instruções. Também pode estruturar comandos formais, como voltar para um ponto específico em um bloco de informação lido ou mesmo suspender o serviço.

O primeiro componente é utilizado para trazer a máquina ao diálogo, que pode ser suprimido conforme o segundo elemento. Verbos como mostrar ou falar, por exemplo, possuem tom imperativo, fato que deve ser relevado no processo. Posteriormente, há uma combinação entre novidades ou informação de fato com a possibilidade de acionar um contexto específico. Regressando ao que foi observado anteriormente, desta forma é possível desenvolver diálogos, não apenas sistemas de estímulo resposta simples, entre agentes de naturezas distintas.

Posteriormente, foi necessário utilizar uma plataforma para a implantação prática do projeto. O estudo da documentação técnica da própria empresa Google levou à definição do uso do serviço *Dialogflow*⁸ (anteriormente conhecido como *Api.ai*), que permite desenvolver produtos com processamento de linguagem natural para uso em assistentes pessoais e *chatbots*. Neste caso, seria apenas uma sequência de respostas para perguntas pré-determinadas através da identificação de palavras-chaves informadas por comandos de voz ou texto. Posteriormente, foi possível combinar informações extraídas de páginas indexadas por sistemas RSS (*Really Simple Syndication*) para automatizar tal processo.

O protótipo foi executado com sucesso, mesmo com algumas dificuldades para a implantação. Entre elas, a vinculação das notícias com as preferências do usuário junto ao Google, fato justificado pela proteção da empresa a informações pessoais. Muitas vezes foi necessário utilizar a leitura em inglês para executar a rotina do diálogo. Meses após a pesquisa, uma voz em português do Brasil foi implantada. No entanto, o trabalho obteve pleno êxito combinando texto em português com a leitura por uma voz nativa em inglês.

Aproximadamente nove meses após a execução do projeto piloto, a empresa Globo.com implantou de forma pública uma aplicação dedicada às eleições (G1, 2018). A experiência relatada nesta pesquisa foi fundamental para tal produto.

Reflexões Finais

Refletir sobre interfaces, considerando as constantes apropriações e mudanças sociais, em diálogo com as táticas dos usuários, é investigar tendo em tela muitas variáveis. Se, como pensa Scolari (2018), anteriormente os aparelhos eram posicionados no centro das investigações, hoje é fundamental compreender todas as nuances que envolvem esta evolução. Além disso, é relevante ir adiante e ocupar, na pesquisa, a posição também da indústria. Chegar à raiz do desenvolvimento dos aparelhos e tentar conhecer e prever como vão se comportar os usuários. E isto ocorreu na investigação aqui descrita.

Os assistentes pessoais que, em muitos países, já ocupam as residências em larga escala, começam com desconfiança de parte do público. Conversar com as máquinas equivale, como citamos, ao

8 Serviço para desenvolvimento de rotinas de interação humano-computador baseadas em conversas em linguagem natural, Recuperado de <https://dialogflow.com>.

imaginário de um filme de ficção. Engloba, de certa forma, mudança de paradigma, em que usuários estão por vezes acostumados com a presença do sentido da visão, que muitas vezes é acompanhada pelo toque. Porém, no caminho da apropriação e, especialmente, da compreensão das vantagens que tal tecnologia pode oferecer, começa o processo de negociação. Para tanto, devemos nos debruçar sobre as áreas de inteligência artificial e *machine learning* pois para serem realmente “*smart*”, os dispositivos devem ter um mecanismo de aprendizado com o interlocutor humano. É uma nova mediação que os estudos em comunicação e design precisam abordar para continuar entendendo e pesquisando este objeto.

O mercado se transforma dentro de novos modelos do ecossistema midiático. Portanto, empresas como a Globo.com precisam buscar formas para que sua atuação continue atingindo o público, que caminha junto a novas tendências. Os resultados da pesquisa se mostraram otimistas, indicando que a prática de assistentes pessoais por voz, no contexto jornalístico, é uma realidade e precisa ser cada vez mais elaborada, de forma a se aproximar do “leitor”, que agora é um usuário ouvinte e interagente de uma interface por voz.

Se, inicialmente, de maneira simplista, se poderia pensar numa conversação a exemplo do rádio e de um retorno às discussões sobre a oralidade, evidencia-se um sistema bem mais complexo. Na perspectiva coevolutiva, todos conversam ao mesmo tempo, em aprendizado constante e mútuo, tendo a voz e a fala em papéis importantes, já que sem elas o homem permaneceria preso apenas às suas ideias. A interface perfeita, porém, é a que permite ao usuário esquecer que ela existe. No momento, isto se cumpre com os assistentes pessoais, pelo uso da voz, inseparável da humanidade desde os mais antigos estágios da consciência.

Referências Bibliográficas

- G1. (2018, setembro 17). G1 responde perguntas sobre eleição para presidente pelo assistente de voz do celular [Blog]. Recuperado de <https://g1.globo.com/economia/midia-marketing/noticia/2018/09/17/g1-responde-perguntas-sobre-eleicao-para-presidente-pelo-assistente-de-voz-do-celular.ghtml> .
- Garun, N. (2017, outubro 7). Botched New York Times promo gave away a Google Home for \$17 [Blog]. Recuperado de <https://www.theverge.com/2017/10/5/16428944/google-home-discount-deal-new-york-times>.
- Hammoudeh, M. & Arioua, M. (2018). *Sensors and actuators in smart cities*. Basel: MDPI.
- Laurel, B. (Ed.). (1990). *Art of Human-Computer Interface Design*. Reading: Addison Wesley.
- Lévy, P. (1999). *Cibercultura*. São Paulo: Editora 34.
- Lotterbach, S. & Peissner M. (2006). Voice User Interfaces in Industrial Environments. In D. T. Pham, E. E. Eldukhri & A. J. Soroka (Eds.), *Intelligent Production Machines and Systems* (pp 592 - 596). UK: Elsevier Science.

- Manovich, L. (2013). *Software Takes Command*. New York: Bloomsbury Academic.
- McLuhan, M. (1964). *Os meios de comunicação como extensões do homem*. São Paulo: Cultrix.
- Mittchell, W. J. (2003). *ME++, The Cyborg Self and the Networked City*. Boston: MIT Press.
- Moggridge, B. (2007). *Designing Interactions*. Cambridge: MIT Press.
- Moorthy, A., Vu, K. (2014). *Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space*. *International Journal of Human-Computer Interaction*. 31(4), 307-335.
- Ong, W. (1998). *Oralidade e cultura escrita*. Campinas: Papirus.
- Pellanda, E. C. (2005). *Internet Móvel: Novas Relações na Cibercultura Derivadas da Mobilidade na Comunicação* (Tese de doutorado). Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brasil.
- Pinker, S. (2004). *O instinto da linguagem: como a mente cria a linguagem*. São Paulo: Martins Fontes.
- Rogers, Y., Sharp, H., & Preece, J. (2013). *Design de Interação: Além da interação humano-computador*. Porto Alegre: Bookman.
- Rugel, A. (2017, setembro 7). Ok Google, Talk to Vogue [Blog]. Recuperado de <https://technology.condenast.com/story/vogue-on-google-home>.
- Saussure, F. (1998). *Course in General Linguistics*. Chicago: Open Court. (Reimpressão da 1ª ed.).
- Scolari, C. (2018). *Las leyes de la interfaz: Diseño, ecología, evolución, tecnología*. Barcelona: Gedisa.
- Kubrick, S (Producer) & Kubrick, S. (Director). (1968). *2001: Uma Odisseia No Espaço* [Filme]. Estados Unidos e Inglaterra: Warner Home Video.
- Vogue. (2017, agosto 17). Go Behind the Scenes With Vogue and Google Home on Interviews With Jennifer Lawrence, Oprah, and More [Blog]. Recuperado de <https://www.vogue.com/article/september-issue-google-home-integration-behind-the-scenes>.