



Departamento de Sociologia

## A Imprensa Online Portuguesa na Época da Web Semântica

João Vitor Gomes Patrício

Dissertação submetida como requisito parcial para a obtenção do grau de

Mestre em Comunicação, Cultura e Tecnologias da Informação

Orientador:

Prof. Doutor António Maria Salvado Coxito Granado – Professor Auxiliar

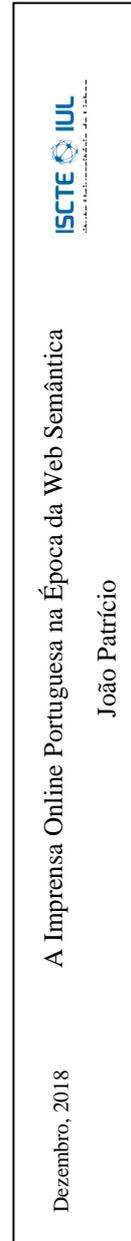
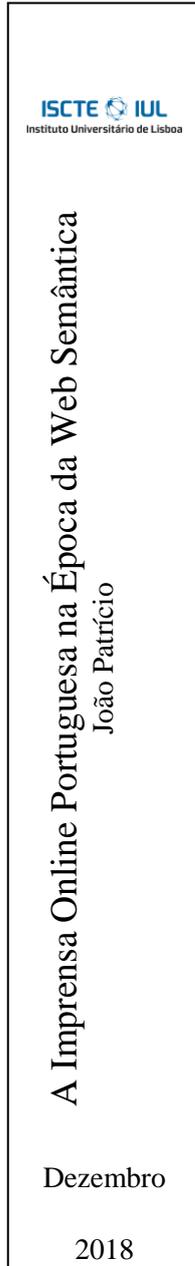
NOVA FCSH-Universidade Nova de Lisboa

Coorientadora:

Mestre Ana Isabel Pinto Martinho – Investigadora Assistente

Centro de Investigação e Estudos de Sociologia-Instituto Universitário de Lisboa

Dezembro 2018



## RESUMO

No avanço da Internet para a uma Web 3.0, as tecnologias ligadas à Web Semântica ganham terreno em diversas aplicações, dando lugar a um novo paradigma online: o da personalização. Olhando para a atividade jornalística, esta nova forma da Internet merece maior atenção, dada a possível influência sobre o consumo de notícias na rede. A presente dissertação pretende fazer uma ponte entre os estudos da Engenharia Informática focados na Web Semântica e os do Jornalismo, de forma a perceber como é que as redações com presença online em Portugal estão a tirar proveito destas tecnologias ou, pelo menos, como poderiam fazê-lo. A investigação conta com o conhecimento de especialistas familiarizados com a Web Semântica no campo académico, bem como de vários quadros de jornais online portugueses, com ou sem conhecimento desta que é tida como “uma extensão” da Web como a concebemos hoje.

**Palavras-chave:** jornalismo; web semântica; comunicação; web 3.0

## ABSTRACT

In the advance of the Internet to a Web 3.0, the technologies connected to the Semantic Web gain ground in several applications, rising to a new online paradigm: customization. Regarding the journalistic activity, this new form of the Internet deserves more attention, given the possible influence on the online news consumption. This dissertation will bridge the studies of Computer Engineering focused on Semantic Web and Journalism, in order to understand how the newspapers with online presence in Portugal are taking advantage of these technologies or, at least, how they could do so. This research counts on the knowledge of experts familiar with Semantic Web from the academic field, as well as several Portuguese online newspapers' personnel, with or without knowledge of what is considered as an "extension" of the Web as we conceive it today.

**Keywords:** journalism; semantic web; communication studies; web 3.0



## ÍNDICE

INTRODUÇÃO .....	1
I - ENQUADRAMENTO TEÓRICO .....	5
1.1 - WEB SEMÂNTICA .....	5
1.2 - A ENTRADA DA WEB SEMÂNTICA NO JORNALISMO.....	8
1.3 - (NOVAS) ROTINAS JORNALÍSTICAS.....	10
1.4 – DO GATEKEEPING PARA O GATEWATCHING: A CAMINHO DE UM TRABALHO DE CURADORIA? .....	14
1.5 – EXEMPLOS DE REDAÇÕES QUE FUNCIONAM COM FERRAMENTAS SEMÂNTICAS .....	15
II - METODOLOGIA .....	17
2.1 – SOBRE OS ENTREVISTADOS .....	19
III – PERCEÇÕES DA WEB SEMÂNTICA .....	22
3.1 – RISCOS E POTENCIALIDADES.....	22
3.2 – CONHECIMENTOS SOBRE WEB SEMÂNTICA NAS REDAÇÕES.....	24
3.3 – PREDISPOSIÇÕES PARA CONHECER E DIVULGAR A WEB SEMÂNTICA NAS REDAÇÕES.....	25
3.4 – COMENTÁRIOS ÀS PROPOSTAS TEÓRICAS.....	26
IV – O CENÁRIO DO JORNALISMO ONLINE PORTUGUÊS.....	30
4.1 – O CASO DO DINHEIRO VIVO/ GRUPO GLOBAL MEDIA.....	30
4.2 – O CASO DO JORNAL DE NEGÓCIOS.....	31
4.3 – O CASO DO OBSEVADOR .....	32
4.4 – O CASO DO PÚBLICO.....	33
4.5 – ARQUIVO DIGITAL RTP: UM CASO DE POTENCIAL APLICAÇÃO DA WEB SEMÂNTICA.....	34
CONCLUSÕES.....	36
POSSÍVEIS PORTAS DE ENTRADA DA WEB SEMÂNTICA.....	38
O QUE PODERIA SER FEITO? .....	39
ÚLTIMAS CONSIDERAÇÕES: ESTUDAR A WEB SEMÂNTICA EM 2018?.....	39
BIBLIOGRAFIA.....	41

ANEXOS.....	I
ANEXO A – ÁREAS DA INTELIGÊNCIA ARTIFICIAL.....	I
ANEXO B – PILHA PROTOCOLAR DE LINGUAGENS QUE SUPORTAM A WEB SEMÂNTICA.....	I
ANEXO C – MINUTA DE ENTREVISTAS PARA INVESTIGADORES.....	II
ANEXO D – MINUTA DE ENTREVISTAS PARA QUADROS DOS JORNAIS.....	III
CURRÍCULO.....	IV

## INTRODUÇÃO

O tema da Web Semântica não é propriamente recente. Trata-se de um conjunto de tecnologias que dão que falar há quase vinte anos, tendo começado pelas palavras de Tim Berners Lee:

“The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” (Berners-Lee *et al*, 2002: 25).

Um dos grandes objetivos da Web Semântica passa pelo desenvolvimento de sistemas capazes de tornar a linguagem humana mais acessível às máquinas. Por meio da sobreposição de linguagens de programação baseadas em HTML, é esperado que o conteúdo presente nas páginas web seja lido e interpretado por computadores, constituindo assim uma nova forma de representar conhecimento. Ora, a Web Semântica apresenta-se, portanto, como apenas uma das várias possibilidades trazidas pelos desenvolvimentos da Inteligência Artificial, tocando em aspetos como a Representação de Conhecimento, o Processamento de Linguagem Natural, Agentes Inteligentes e o *Machine Learning*. Monard e Baranauskas (2000) apresentam uma possível tipologia das várias áreas afetas à Inteligência Artificial, da qual fazem justamente parte o *Machine Learning* (vulgo “Aprendizagem da Máquina”), e o Processamento da Linguagem Natural (cf. Anexo A).

O novo estágio da Web, a Web 3.0, passa por uma maior aproximação das máquinas no processo de criação de conteúdo na rede, entre o produtor e o consumidor: “Web 3.0 has changed the entire process by bringing machines closer to the users and producers for more dynamic, interactive and efficient creation of the contents and its management” (Pattal *et al*, 2009: 126). Assente num paradigma de personalização, a nova Web permite que o utilizador aceda primeiramente e com maior facilidade a conteúdo baseado no seu histórico de navegação, histórico de localização, conteúdos publicados nas redes sociais, entre outros.

A Web Semântica não é uma realidade assim tão remota quanto isso. Rafael Reis (2016) refere que aplicações como o Google Now ou o Google Search Engine já integram estas tecnologias, “onde se verifica uma capacidade de pesquisa e apresentação de informação baseada nos gostos e preferências do utilizador (*ibid*, 2016:12). Estas inovações podem parecer entusiasmantes, mas se direcionarmos estes avanços para a ótica da produção e consumo noticiosos, os ânimos poderão alterar-se.

Tendo como questão de partida “Qual a influência da Web Semântica na atividade da imprensa online portuguesa?”, esta dissertação pretende, primeiramente, (re)construir uma ponte teórica entre as ciências da comunicação e a engenharia informática. A temática do Jornalismo e da sua relação com a Web Semântica mereceu alguma atenção em termos de produção académica ao longo da primeira década do século XXI, mas a partir de 2010 nota-se uma quebra na discussão. José Nuno Pereira, impulsor da Internet no contexto português, afirma que a Web Semântica é o resultado de um contexto complexo,

marcado por uma comunicação social “estrangulada em termos de dinheiro”, nas mãos de canais como o Facebook e o Google, pelo crescimento do consumo de conteúdo a partir de dispositivos móveis baseado em linguagens de programação onde não há interesse na Web Semântica, mas sim na rentabilidade financeira:

“...acho que o problema da Web Semântica é que não deixa de interessar a quem produz os conteúdos, mas para quem produz os conteúdos, genericamente, a WS não é uma prioridade, porque têm de sobreviver. Portanto acho que não é que não continue a haver interesse, mas o interesse pela Web Semântica foi atropelado pela loucura da eficácia, de chegar mais rápido. Quem poderia ter salvaguardado isso, que seriam o Google e o Facebook, eu acho que, aparentemente, não deverão ter muito interesse e, portanto, as coisas ficam penduradas.” (José Nuno Pereira, 4 de junho de 2018).

Fazendo ressurgir a discussão, a presente dissertação partirá de uma reflexão teórica, e fará um paralelo com o plano prático e empírico do Jornalismo. Esta investigação percorrerá vários jornais online portugueses com o intuito de perceber se as tecnologias semânticas já se fazem sentir nas rotinas de trabalho ou se, no mínimo, há conhecimento da existência de tais inovações dentro das redações.

Rafael Reis conclui na sua tese de mestrado que a relação entre o jornalismo português e a Web 3.0 “é ainda muito distante”, dado que muitos profissionais na área desconhecem tal estágio da Web e que se verifica a “ausência de uma aposta na Web Semântica” (*ibid*, 2016: 124). Neste estudo pretende-se dissociar os termos “Web Semântica” e “Web 3.0”. Adotando a visão de Patal *et al* (2009), a Web Semântica é apenas uma parte de todo um fenómeno que é a Web 3.0, apenas a vertente tecnológica, aquela que já está em funcionamento e pode ser efetivamente estudada. A Web 3.0 comporta, para além desta componente tecnológica, um ambiente de computação social (*ibid*, 2009: 126). Este trabalho procura ainda dar uma continuidade a este estudo, alargando o espetro de jornais em estudo, encontrar quem tem conhecimentos sobre a Web Semântica nos media portugueses, e a montante, compreender as perceções deste grupo de tecnologias.

A presente investigação tem como objetivo avaliar se, passados dois anos após a dissertação de Rafael Reis, poderá haver vestígios de Web Semântica no jornalismo online português; ou, em caso negativo, perceber o que está a ser e o que poderia ser feito. Pensar atualmente na Comunicação é pensar à luz do tempo presente, e a atualidade da Comunicação passa impreterivelmente pelas condições técnicas e tecnológicas em que se desenrola. Junto de quem tem conhecimentos sobre a Web Semântica, e de quem trabalha em jornais portugueses com presença online, tenciona-se perceber o que é de facto esta Web Semântica e a sua base tecnológica adjacente, que riscos e potencialidades trazem para o setor do Jornalismo, que redações estão abertas para este “novo” panorama tecnológico, e tentar chegar a formas de dar o passo em frente.

Nesta dissertação pretende-se ainda confrontar o plano teórico com o plano prático, tentando perceber como é que os autores concebem a ligação entre estas duas áreas de conhecimento, e como é que os profissionais do Jornalismo respondem às suas propostas.

Numa primeira fase foram entrevistados especialistas na área da Engenharia Informática, de forma a complementar a literatura já escrita na área sobre a Web Semântica. Foram também entrevistados vários editores, diretores e jornalistas portugueses, de modo a compreender como é que a Web Semântica se liga ao Jornalismo, bem como para traçar o panorama nacional. Estas entrevistas tiveram também como objetivo perceber se as tecnologias semânticas são uma realidade assim tão distante das redações portuguesas.

Mais do que produzir ou co-produzir conhecimento, fazer ciência também deve ser uma forma de gerar ação. Esta dissertação não se circunscreve num documento escrito. A escolha deste tema vem no sentido de criar *awareness* para o poder que as tecnologias semânticas podem ter para o Jornalismo. Num panorama marcado por expressões como “fake news”, “echo chambers”, “filter bubbles”, “algoritmos”, “big data”, “influencers”, entre muitas outras, estudar o Jornalismo à luz dos mais recentes avanços tecnológicos nunca fez tanto sentido.

Estudar a Web Semântica no campo do Jornalismo significa diagnosticar um dos pilares-base da nossa democracia; e falar de uma filtragem no acesso ao conteúdo noticioso é falar de um problema que poderá comportar sérias consequências a curto e médio prazo. Eli Pariser (2011) apresenta na obra “The Filter Bubble” uma certa visão de desencantamento com os recentes avanços da Internet, nomeadamente com a sua capacidade de atomizar os indivíduos por esferas de interesses:

“For a time, it seemed that the Internet was going to entirely redemocratize society. Bloggers and citizen journalists would single-handedly rebuild the public media. Politicians would be able to run only with a broad base of support from small, everyday donors. Local governments would become more transparent and accountable to their citizens. And yet the era of civic connection I dreamed about hasn’t come. Democracy requires citizens to see things from one another’s point of view, but instead we’re more and more enclosed in our own bubbles. Democracy requires a reliance on shared facts; instead we’re being offered parallel but separate universes.” (*ibid*, 2011: 8).

Neste sentido, o maior desafio apontado por este autor é o poder de distorção dos filtros junto dos utilizadores da Internet: “personalized filters limit what we are exposed to and therefore affect the way we think and learn” (*ibid*, 2011: 48). O viés da confirmação é apenas uma das consequências. Sermos confrontados com as nossas próprias ideias e perspetivas do mundo é muito mais fácil e confortável do que nos depararmos com o antagónico, o que é diferente, o que nos desafia (*ibid*, 2011: 51).

Numa altura em que tantos apregoam “a morte do Jornalismo”, este nunca foi tão necessário como agora. Numa atualidade em que as fontes de informação se multiplicam e o paradigma da imediatez pressiona as redações, o tato jornalístico assume uma importância considerável. Numa Web onde a personalização progride de forma invisível aos olhos dos utilizadores e fora do nosso controlo (*ibid*, 2011: 119), é crucial olhar de forma crítica para o progresso tecnológico. Independentemente do panorama que o jornalismo português esteja a atravessar frente à Web Semântica, esta dissertação vem no sentido de dar a conhecer (novos) pontos de interesse e de inovação dentro do jornalismo.

A presente dissertação partirá com as seguintes hipóteses de investigação:

1. A Web Semântica continua a ser um tema que ainda muito poucos conhecem nas redações portuguesas.
2. Há um desfasamento entre o plano teórico e o plano prático no que toca à introdução das tecnologias ligadas à Web Semântica nas redações.
3. O jornalista enquanto gatekeeper poderá ceder lugar a um jornalista enquanto gatwatcher; passando a ter uma função de curadoria da informação.
4. Num contexto marcado pela influência crescente da curadoria e/ou filtragem do conteúdo por parte de algoritmos (primeiramente desencadeada pelos próprios utilizadores), o jornalista atua como “nutricionista” da dieta mediática dos leitores.
5. A entrada da Web Semântica passa por uma maior colaboração entre jornalistas e engenheiros informáticos, em contexto de redação.
6. As novas exigências para o jornalista guerreiam com o paradigma da imediatez que se vive nos órgãos de comunicação online.

# I - ENQUADRAMENTO TEÓRICO

## 1.1 - WEB SEMÂNTICA

A definição mais corrente de Web Semântica é dada pelo seu próprio criador, Tim Berners-Lee. Após a viragem do século, Berners-Lee *et al* (2002) referem esta nova Web como sendo uma extensão da atual, na qual é dada à informação um significado bem definido, permitindo uma maior cooperação entre computadores e pessoas (*ibid*, 2002: 25). A partir daqui a Web avançará em breve para o seu terceiro estágio. Para Pattal *et al* (2009), tratar-se-á de aproximar máquinas e seres humanos, fornecendo aos utilizadores a informação que querem, quando quiserem (*ibid*, 2009: 126).

Apesar da grande referência à Web Semântica de Tim Berners-Lee de 2002 ser a mais citada entre vários estudos na área, esta nova fase da Web é algo que tem vindo a fazer parte dos planos do criador da World Wide Web desde sempre, e uma das primeiras menções surge na obra “Weaving The Web”:

“I have a dream for the Web... and it has two parts. (...) In the second part of the dream, collaborations extend to computers. Machines become capable of analyzing all the data on the Web – the content, links and transactions between people and computers. A “Semantic Web”, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives will be handled by machines talking to machines.” (Berners-Lee, 2000: 157-158).

Um conceito transversal neste campo é o de “agente”, enquanto o intermediário que intercede entre o utilizador e os ficheiros na World Wide Web. Grigoris Antoniou e Frank van Harmelen (2008) definem estes elementos como “pieces of software that work autonomously and proactively. Conceptually they evolved out of the concepts of object-oriented programming and component-based software development” (*ibid*, 2008: 15). Tim Berners-Lee refere que tais agentes constituem uma peça chave no desenvolvimento da Web Semântica. Estes são definidos como programas ou pedaços de software que funcionam sem controlo ou supervisão constante de um ser humano, e que colecionam conteúdo da Web a partir de várias fontes, bem como processam informação e trocam resultados com outros programas (Berners-Lee *et al*, 2002: 27-28). Estes agentes têm a sua performance controlada a partir de outros programas específicos, os centros de reputação. Estes determinam se os agentes procederam bem ou mal, criando-lhes, precisamente, um rótulo de reputação, que influenciará processos de, por exemplo, procura de serviços no futuro; até porque os agentes inteligentes não funcionam necessariamente isolados, mas antes em cadeia. Um circuito, a título de exemplo, pode ser formado por um agente de procura de informação, outro de compilação, e um último de sumarização.

A Web Semântica funciona a partir de um conjunto de camadas, camadas essas compostas de linguagens de programação que permitem que os dados se tornem mais perceptíveis para os computadores e não

apenas para leitores humanos (cf. Anexo B). Doravante o modelo será chamado de “pilha protocolar”, um termo introduzido por Vítor Basto Fernandes em contexto de entrevista. Nas camadas mais baixas, encontram-se os formatos de representação de dados mais simples, aos quais se vão acrescentando camadas de significação e de relação.

Na base deste esquema está a XML (*eXtensible Markup Language*). À semelhança da HTML (*Hypertext Markup Language*), trata-se de uma linguagem de marcação na qual são escritas as páginas web. Contudo, a XML permite a definição de *tags* por parte do utilizador, de forma a identificar os vários elementos presentes na página, ao invés da identificação estandardizada por cabeçalhos (*headers*) e corpo (*body*) (Berners-Lee *et al*, 2002: 27). Ambas as linguagens permitem criar conteúdos e dar informações sobre o papel que cada pedaço de conteúdo representa na página, os chamados metadados, e ambas permitem uma leitura relativamente fácil para seres humanos. No entanto, os documentos HTML não oferecem informações relativas aos seus vários componentes e das relações entre si; já em documentos XML a informação torna-se um pouco mais acessível às máquinas, uma vez que cada pedaço de informação é descrito (Antonίου e van Harmelen, 2008: 26). Ainda neste nível encontra-se a *XML Schema*, uma linguagem de esquema baseada em XML, e que permite uma leitura mais fácil dos documentos XML (*ibid*, 2008: 39). Os esquemas, segundo Vítor Basto Fernandes, permitem representar relações ou recursos.

Apesar da universalidade da linguagem XML, esta não providencia quaisquer informações semânticas, isto é, relacionadas com os significados, relativas aos dados presentes nos documentos da Web (*ibid*, 2008: 65). É aqui que surge o nível da RDF (*Resource Description Framework*), um modelo de dados que fornece sintaxe à XML, num tríptico de objeto-atributo-valor, formando assim declarações (*statements*) (*ibid*, 2008: 66).

À semelhança da XML, a RDF também se faz acompanhar de uma linguagem de esquema, a RDF Schema ou RDFS. À medida que a RDF permite aos utilizadores descrever os recursos (objetos, categorias...) através do seu próprio vocabulário, é a partir da RDFS que podem ser feitas considerações e definições das semânticas entre termos (*ibid*, 2008: 84) (cf *ibid*, 2008: 88).

“XML Schema constrains the structure of XML documents, whereas RDF Schema defines the vocabulary used in RDF data models. In RDFS we can define the vocabulary, specify which properties apply to which kinds of objects and what values they can take, and describe the relationships between objects.”  
(*ibid*, 2008: 66).

Enquanto a RDF trata de declarações, compostas pela já referida tripla recurso-propriedade-valor, a RDFS trata relações de classe-subclasse, relações de propriedade-subpropriedade, e de restrições de domínios e gamas (*ibid*, 2008: 109).

Ainda assim, A RDF e a RDFS oferecem uma expressividade muito limitada: “RDF is (roughly) limited to binary ground predicates, and RDF Schema is (roughly) limited to a subclass hierarchy and a property hierarchy, with domain and range definitions of these properties.” (*ibid*, 2008: 113). Chega-se aqui ao nível da ontologia, marcado por um conjunto de linguagens que permitem aos utilizadores escrever conceptualizações de modelos de domínios de forma explícita e formal. São linguagens com uma sintaxe bem definida, um suporte de raciocínio eficiente, uma semântica formal e um maior poder e conveniência expressivos, de forma a facilitar o processamento da informação pelas máquinas (*ibid*, 2008: 114). A OWL (*Web Ontology Language*) é a maior referência a este nível. Segundo Pattal *et al*, RDF e OWL são praticamente a mesma coisa, no entanto, esta última consiste numa linguagem mais forte, com um vocabulário mais alargado e uma sintaxe também ela mais forte, e garante uma maior “interpretabilidade” entre computadores quando comparada com a primeira (Pattal *et al*, 2009: 126). A este nível surge a lógica de predicados, onde mais do que relações entre recursos, são feitas inferências sobre essas mesmas relações.

Na camada acima encontra-se o nível da Lógica (*Logic*), destinada a melhorar o trabalho das linguagens ontológicas no que concerne à criação de conhecimento declarativo (Antoniou e van Harmelen, 2008: 18-19), isto é, descrevendo a informação sem ter em conta a forma como pode ser deduzida (*ibid*, 2008: 13).

O processo dedutivo *per se* encontra-se ao nível da Prova (*Proof*), numa camada onde pertencem igualmente as ações de representação e de validação de provas em linguagens Web pertencentes a camadas mais baixas (*ibid*, 2008: 18).

No topo encontra-se o nível de Confiança (*Trust*), onde as assinaturas digitais são particularmente importantes, e a informação é tida como fidedigna por meio de agentes de confiança, agências de certificação e rating e organizações de consumidores. Para os autores, este conceito de confiança é de grande importância, pois defendem que a Web apenas atingirá o seu pleno potencial quando os utilizadores tiverem confiança nas suas operações e na qualidade da informação disponibilizada (*ibid*, 2008: 18).

## 1.2 - A ENTRADA DA WEB SEMÂNTICA NO JORNALISMO

As tecnologias semânticas necessitam, por enquanto, de um contributo humano no seu funcionamento, e a área jornalística é exemplo disso mesmo. Para Corrêa e Bertocchi (2012), trata-se de um momento ideal para repensar a posição do jornalista face a estas novas tecnologias:

“Na nossa visão, uma WS socialmente relevante não irá falar por si mesma, mas por meio da Comunicação. Ou seja, sem o comunicador como agenciador dessa rede – ao lado da máquina – há somente uma web semântica totalitária. Cabe, portanto, repensar o posicionamento do jornalista e das empresas informativas diante deste próximo panorama.” (*ibid*, 2012: 125).

Ambas as autoras veem o jornalista como um profissional que, no meio do *information overload*, assume o papel de especialista, organizando e dando sentido aos dados (*ibid*, 2012: 125). Trata-se de um contexto onde a abundância de informação, de fontes e de narrativas coloca em risco a transformação dos dados em informação e em conhecimento (*ibid*, 2012: 131). Embora o poder de processamento de informação por parte dos computadores esteja a tornar-se cada vez mais rápido e melhor, podendo até, no futuro, criar um conhecimento artificial, as autoras defendem que o jornalista permanece indispensável por dois motivos. Primeiramente, ainda nenhum robot passou no teste de Turing, e, portanto, os sistemas de algoritmos ainda não conseguem entrar em diálogo com a sua audiência; em segundo lugar, as tecnologias semânticas, ao alimentarem os consumidores com as informações que desejam, num regime de personalização máxima, criam aquilo que Eli Pariser apelida de “filter bubbles” (*ibid*, 2011: 136). Pariser (2011) define estas bolhas como um universo temático personalizado para cada utilizador, criado a partir de sistemas que pretendem prever quem é o utilizador e o que poderá ser do seu interesse:

“The basic code at the heart of the new Internet is pretty simple. The new generation of Internet filters looks at the things you seem to like—the actual things you’ve done, or the things people like you like—and tries to extrapolate. They are prediction engines, constantly creating and refining a theory of who you are and what you’ll do and want next. Together, these engines create a unique universe of information for each of us—what I’ve come to call a filter bubble—which fundamentally alters the way we encounter ideas and information.” (*ibid*, 2011: 10).

Para Pariser, estas bolhas assentam em três dinâmicas. Primeiramente, o indivíduo encontra-se só no interior da sua bolha temática (*ibid*, 2011: 10). Em segundo lugar, a bolha é invisível: “Google doesn’t tell you who it thinks you are or why it’s showing you the results you’re seeing”; ou tampouco sabemos se as constatações que os motores de busca fazem sobre nós são verdadeiras ou até se são mesmo sobre nós (*ibid*, 2011: 10). Por fim, o indivíduo não escolhe entrar na bolha; enquanto optamos por determinados canais de televisão ou estações de rádio definimos conscientemente o nosso filtro para

fazer sentido do mundo que nos rodeia, o mesmo processo de escolha não é feito com os filtros na Internet. “They come to you—and because they drive up profits for the Web sites that use them, they’ll become harder and harder to avoid.” (*ibid*, 2011: 11).

*Flaxman et al* (2016), afirmam que, efetivamente, com a grande oferta de fontes de informação, os consumidores escolham aquelas que vão ao encontro daquilo que defendem e acreditam, criando assim as chamadas “echo chambers” (Sustein, 2009 *apud ibid*, 2016: 299). Ainda, motores de busca, agregadores e redes sociais online, através de modelos de aprendizagem automática fomentam o surgimento das “filter bubbles”, “in which algorithms inadvertently amplify ideological segregation by automatically recommending content an individual is likely to agree with.” (*ibid*, 2016: 299). Contudo, sustentam os mesmos autores, o surgimento de plataformas como as redes sociais online contribui para uma maior exposição a uma diversidade de ideias, libertando os consumidores de um “padrão de consumo insular”<sup>1</sup> (Benkler, 2006; Obendorf *et al.* 2007; Goel, Hofman, e Siner 2012 *apud ibid* 2016: 299). Apoiando-se numa bateria de outros autores, *Flaxman et al* argumentam que as interações sociais aumentam a exposição a perspectivas mais heterogéneas, e que nas redes sociais online, uma considerável quota das relações é feita entre indivíduos de diferentes posições do espectro político (*ibid*, 2016: 299).

A partir de uma amostra composta pelos padrões de pesquisa de 50.000 internautas localizados nos Estados Unidos da América que leem notícias com regularidade, os mesmos autores afirmam que a exposição a perspectivas antagónicas é mais alta em canais associados a maiores graus de segregação, e que as redes sociais online e os motores de busca não transformaram o consumo de notícias ao nível que se previa (*ibid*, 2016: 299-301).

Olhando para a ideia de Pariser de que a personalização nos motores de busca leva à criação das já referidas bolhas temáticas, os autores apresentam duas possíveis explicações: A primeira tem a ver com o facto dos utilizadores influenciarem os sistemas de aprendizagem automática dos resultados nos motores de busca através da sua formulação terminológica nas pesquisas, como pesquisar por “obamacare” em vez de “health care reform” (Borra e Weber, 2012 *apud ibid*, 2016: 311-312); a segunda explicação apresentada tem a ver com o facto de, mesmo perante os mesmos resultados das pesquisas, os utilizadores vão mais provavelmente seleccionar aqueles que partilham a mesma ideologia política (Garrett, 2009; Iyengar e Hahn, 2009; Munson e Resnick, 2010 *apud ibid*, 2016: 312).

De forma sumária, uma das conclusões destes autores centra-se no seguinte: “though the predicted filter bubble and echo chamber mechanisms do appear to increase online segregation, their overall effects at this time are somewhat limited” (*ibid*, 2016: 313).

Dados mais recentes, recolhidos no Digital News Report 2018 do Reuters Institute for the Study of Journalism apontam que 53% dos utilizadores da Internet inquiridos preferem que algoritmos lhes

---

<sup>1</sup> Tradução livre.

sugiram que conteúdos noticiosos consumir, nomeadamente através de redes sociais, motores de busca e agregadores de notícias, contra 44% que preferem que tal trabalho seja feito por editores ou jornalistas (Newman *et al*, 2018: 14). No cenário português, o mesmo estudo aponta que 48% dos inquiridos confia nas notícias encontradas a partir de motores de busca, e 29% confia nos conteúdos noticiosos apresentados nas redes sociais; trata-se de uma redução face ao ano passado, o que, segundo os autores, “indicate concern over the reliability of news in third-party environments” (Martinho, Cardoso e Paisana, 2018, *apud ibid*, 2018: 95-96).

O mesmo estudo também coloca a questão se os algoritmos estreitam ou alargam o espectro temático de conteúdos dos internautas. Os dados relativos a 2017 revelam que estes sistemas expõem a maioria dos utilizadores a um maior conjunto de fontes de informação. Prova disso está na proporção de pessoas que concorda que são expostos a fontes a que normalmente recorreriam. Seja em motores de busca, redes sociais online ou agregadores, a maioria dos inquiridos concorda que está exposta a mais fontes de informação (Newman *et al*, 2017: 16). Em Portugal, os dois últimos relatórios apontam para o surgimento de novas fontes de informação, nomeadamente o Jornal Económico e o ECO em 2017 (Martinho e Cardoso, 2017, *apud ibid*, 2017: 85).

Um outro aspeto levantado pelo Digital News Report tem a ver com a influência dos próprios utilizadores na forma como estes podem controlar e personalizar a sua experiência e acesso à informação nos três canais acima referidos:

“In social media almost half of our respondents (45%) have actively managed the content of their feeds by adding or removing users, with around a third (30%) fine-tuning content in other ways, such as hiding content from particular users. This activity allows users to protect themselves from views they don’t like, but it also opens up the possibility that they just end up seeing more of what they like.” (*ibid*, 2017: 16).

### **1.3 - (NOVAS) ROTINAS JORNALÍSTICAS**

Em todas as redações, as rotinas de produção têm o mesmo denominador comum: “a escassez de tempo e de meios” (Wolf, 2006: 218). Segundo o mesmo autor, o processo de criação do produto final que é a notícia à qual o leitor ou espectador têm acesso passa por três fases. São elas a recolha, a seleção e a apresentação da informação (*ibid*, 2006: 218).

A fase da recolha consiste na obtenção de informações que poderão ser ou não notícia, de acordo com um conjunto daquilo a que a teoria do *newsmaking* apelida de “valores-notícia” (cf. Wolf, 2006: 195). Neste primeiro momento, as fontes desempenham um importante papel ao providenciarem “material informativo já facilmente inserível nos procedimentos produtivos normais da redacção”. Um dos

exemplos dados é o das agências, que procedem a um “corte” do fluxo de acontecimentos, e o trabalho do jornalista, aquando da recolha, consiste num “ajustamento marginal” (Wolf, 2006: 220).

Num momento de seleção da informação, o material reunido é convertido na forma de notícia. Trata-se de um “processo complexo, que se desenrola ao longo de todo o ciclo de trabalho, realizado a instâncias diferentes – desde as fontes até ao simples redactor – e com motivações que não são todas imediatamente imputáveis à necessidade directa de escolher as notícias a transmitir” (*ibid*, 2006: 241). Relevância e eficiência (em termos de pessoal, formato e tempo de produção) são os grandes critérios na seleção das notícias apontados pelo mesmo autor (*ibid*, 2006: 241). É aqui que entra a função de *gatekeeping*, que será explicada na secção seguinte.

Por fim, a apresentação das notícias, a fase onde a informação, depois de recolhida e tratada, chega aos consumidores de informação:

“...se todas as fases anteriores funcionam no sentido de descontextualizar os factos do quadro social, histórico, económico, político e cultural em que acontecem e em que são interpretáveis (isto é, no sentido de «curvar» os acontecimentos às exigências de organização do trabalho informativo), nesta última fase produtiva, executa-se uma operação inversa: recontextualizam-se esses acontecimentos, mas num quadro diferente, dentro do formato do noticiário.” (*ibid*, 2006: 244).

De acordo com Daniela Bertocchi (2010) as tecnologias semânticas podem afetar as redações a vários níveis neste processo de produção, a jusante do trabalho jornalístico:

“Considerando especificamente as narrativas jornalísticas, as tecnologias da WS interferem em pelo menos cinco níveis da informação: na entrega informativa (como os dados chegam aos usuários, em quais dispositivos e com qual aparência); na pesquisa de dados (como as notícias são recuperadas pelos usuários); na exploração e visualização dos dados (como as informações são visualmente apresentadas aos usuários) e, ainda, na forma de percepção do texto (como as notícias são compreendidas pelos usuários).” (*ibid*, 2010: 8).

### **1.3.1 – O JORNALISTA ENQUANTO CURADOR**

Numa quase “ditadura do algoritmo”, em que estes sistemas “vão decidir o que iremos assistir, ler, ouvir, saber, conforme aquilo que julgarmos relevantes, mas não necessariamente consoante aquilo que for relevante do ponto de vista de toda uma sociedade” (Corrêa e Bertocchi, 2012: 137), o jornalista adquire um papel de curador, o elemento de intervenção humana que trabalha lado a lado com o algoritmo,

mostrando aos consumidores não só aquilo que querem ver, como aquilo que também lhes pode ser desconfortável:

“Um curador de conteúdos é capaz de agregar novas e inusitadas perspectivas à informação, oferecendo aos seus usuários a surpresa, o inesperado ou simplesmente aquilo que o usuário nem imaginaria existir no mundo e sobre o mundo, ampliando seu próprio entendimento de mundo.” (*ibid*, 2012: 137).

Simon Rogers (2011) dá o exemplo do jornalismo de dados como uma área do Jornalismo que adquire este estatuto de curadoria (Rogers, 2011: loc. 56-57 *apud* Corrêa e Bertocchi, 2012: 140). É justamente na abundância de informação em circulação que, afirma o autor, “encontrar a informação correta pode se transformar numa atividade jornalística tão intensa quanto buscar os melhores entrevistados para uma matéria. Trata-se a este nível de olhar para os dados e tentar retirar aquilo que é interessante e novo, assim como tentar cruzamentos de dados para obter factos interessantes (*ibid* 2011: loc. 56-71, *apud* Corrêa e Bertocchi, 2012: 140<sup>2</sup>).

### 1.3.2 - O JORNALISMO SEMÂNTICO SOCIAL

Outra proposta apresentada passa pelo desenvolvimento de um “jornalismo semântico social”. Heravi e McGinnis (2015) apontam que a omnipresença de novas tecnologias tem vindo a permitir que indivíduos e comunidades adquiram o estatuto de fontes primárias de informação e que, portanto, esse jornalismo de cidadão pode ser uma fonte de informação valiosa para órgãos de comunicação social em todo o mundo (*ibid*, 2015: 1).

As autoras desenham um panorama atual do jornalismo orientado por prazos apertados, onde o processamento de volumes consideráveis de conteúdo gerado por utilizadores que chega a ritmos cada vez mais acelerados não pode ser feito pelos métodos tradicionais (*ibid*, 2015: 1-2). Neste sentido, o jornalismo semântico social vem trazer uma solução que pretende ligar todo o conteúdo criado por utilizadores por toda a Web para uma base de dados para efeitos de verificação da veracidade dos factos (*ibid*, 2015: 3). Para a sua concretização, as autoras referem duas ontologias-chave: a SIOC (*Semantically Interlinked Online Communities*), que interliga várias plataformas sociais, devolvendo à Web informações sobre comunidades online em RDF; e a FOAF (*Friend Of a Friend*), uma ontologia que descreve pessoas e as interligações entre si (*ibid*, 2015: 3-4).

O jornalismo semântico social assenta em quatro pilares:

---

<sup>2</sup> Tradução das autoras.

- *Content Discovery*, que passa pela integração da informação em bruto e pelo seu subsequente enriquecimento com meta dados semânticos, de forma a que possa ser utilizada posteriormente (*ibid*, 2015: 5);
- *Filtering and Contextualisation*, onde os metadados da fase anterior são refinados e relacionados com outros conteúdos, colocando as notícias num contexto mais alargado. É aqui que começa a ser traçada a proveniência dos factos e é verificado o seu valor-notícia para que siga para junto dos jornalistas (*ibid*, 2015: 5);
- *Trust Verification*, numa fase em que são utilizadas as informações relativas à proveniência dos factos e aos metadados recolhidos para verificar e detetar a sua veracidade, credibilidade e transparência (*ibid*, 2015: 6);
- *Publication*, um momento dedicado à anotação, publicação e arquivação das peças produzidas, alimentando novamente a primeira fase do processo para a produção de novas notícias (*ibid*, 2015: 6).

### 1.3.3. – O JORNALISTA ENQUANTO “NUTRICIONISTA” DA DIETA MEDIÁTICA

Por fim, uma última proposta poderá ser desenhada a partir da ideia de dieta mediática. Pariser explica que, tal como o corpo do ser humano está programado para consumir açúcares e gorduras, também está programado para server conteúdos que o estimulem, seja violento, de cariz sexual, ofensivo ou embaraçoso. Da mesma forma que o sistema agrícola molda a forma como comemos, os media moldam a informação que consumimos (Pariser, 2011: 13). Num contexto em que o consumo de conteúdo mediático está a passar por um paradigma de personalização, circunscrevemo-nos num espaço de “autopropaganda invisível”, “doutrinando-nos nas nossas próprias ideias amplificando o nosso desejo por aquilo que é familiar, deixando-nos absortos dos perigos à espreita no território escuro do desconhecido”<sup>3</sup> (*ibid*, 2011: 13).

Olhando para o passado, Pariser afirma que os media levavam uma promessa de representatividade dos temas que compunham a atualidade. Na era das filter bubbles, os leitores já não veem aquilo que à partida não lhes interessa, ou tampouco têm a noção de que há enormes acontecimentos e eventos no mundo do qual não estão a ter conhecimento. “We’re never able to experience the whole world at once. But the best information tools give us a sense of where we stand in it” (*ibid*, 2011: 60-61).

Neste sentido urge a necessidade de questionar se os jornalistas têm ou não um papel de “nutricionistas” na nossa dieta mediática.

---

<sup>3</sup> Tradução livre.

## 1.4 – DO GATEKEEPING PARA O GATEWATCHING: A CAMINHO DE UM TRABALHO DE CURADORIA?

No fluxo de comunicação, o jornalista, em conjunto com toda uma equipa editorial, tem recebido o tradicional papel de *gatekeeper*. O termo foi primeiramente reconhecido por Kurt Lewin (1947). No artigo “Channels of Group Life”, o psicólogo explica a dinâmica necessária para proceder a alterações sociais. Para tal, o autor sublinha a necessidade de encontrar indivíduos que se encontrem em “posições chave” (*ibid*, 1947: 143). Nos vários canais de comunicação pelos quais se movem as peças noticiosas existem secções com portões governadas por regras imparciais ou por *gatekeepers*, definidos como “an individual or group [that] is ‘in power’ for making the decision between ‘in’ or ‘out’” (*ibid*, 1947: 145).

A partir deste, David Manning White (1964) traça dois pontos de conclusão relevantes: “the flow of any news items would be through certain channels and, more important, that certain places within these channels would serve as ‘gates’ through which given news items might or might be admitted” (*ibid*, 1964: 160). Pegando num exemplo prático (neste caso foi uma audição no Senado sobre um novo imposto à Educação), o autor refere que o primeiro “portão” na produção noticiosa é o próprio correspondente no local, que decide se a história é “importante” ou não (*ibid*, 1964: 162-163).

Até chegar ao jornal, as notícias passam por uma série de portões: “From reporter to rewrite man, through bureau chief to ‘state’ file editors at various press association offices, the process of choosing and discarding is continuously taking place”, até chegar ao editor-chefe, o alvo principal do estudo de White (*ibid*, 1964: 163). É neste posto que é decidido que notícias seguem para a edição do jornal. Para o autor, a comunicação das notícias é um processo “altamente subjetivo”, baseado nos juízos de valor, nas experiências, atitudes e expectativas destes indivíduos (*ibid*, 1964: 165). Na visão de G. Robinson (1981) as decisões tomadas pelo *gatekeeper* são-no “a partir de uma avaliação individual da noticiabilidade do que em relação a um conjunto de valores que incluem critérios, quer profissionais, quer organizativos, tais como a eficiência, a produção de notícias, a rapidez” (*ibid*, 1981: 97 *apud* Wolf, 2006: 181). Donohue, Tichenor e Olien (1972) falam desta prática como um conjunto de “formas de controlo da informação, que podem estabelecer-se nas decisões acerca da codificação das mensagens, da selecção, da formação da mensagem, da difusão, da programação, da exclusão de toda a mensagem ou das suas componentes” (*ibid*, 1972: 43 *apud* Wolf, 2006: 181-182). Com efeito, segundo Warren Breed (1955), o jornalista “redefine os seus próprios valores ao nível mais pragmático do grupo redactorial”<sup>4</sup> (*ibid*, 1955: 335).

Contudo, com o decorrer do tempo, o papel de *gatekeeping* tem vindo a conhecer algumas ameaças. Avançando para a década de 2000, com a emergência da Web 2.0, surge um novo paradigma de navegar na Internet, um modelo de interação entre produtores e consumidores de informação. É nesta altura que

---

<sup>4</sup> Tradução Livre.

se fala do surgimento dos blogues e das redes sociais online, onde o utilizador já não é um mero consumidor de conteúdo, mas também um produtor, ou melhor, nas palavras de Alvin Tofler ou de Henry Jenkins, um *prosumer*. Para Bruns (2014), as práticas de gatekeeping foram fruto de uma época marcada pela “escassez de canais de notícias e de espaço para as notícias nesses canais” (*ibid*, 2014: 228). Agora, com a multiplicação de plataformas sociais online e a pressão trazida por canais de notícias ativos 24 horas, o autor aponta para uma alteração “na direção do *gatewatching*”, numa cooperação com as audiências (*ibid*, 2014: 226). Bardoel e Deuze corroboram esta dinâmica, relegando ao jornalista um papel de “orientador”, num contexto marcado pelo “aumento explosivo das informações em uma escala mundial, [onde] a necessidade de oferecer informações sobre informações se tornou uma adição crucial às habilidades e tarefas do jornalismo” (Bardoel e Deuze, 2001: 94 *apud* Bruns, 2014: 229).

Com a chegada da Web Semântica, o processo de produção noticioso vê-se a braços com um terceiro *player*: os computadores e os algoritmos. Com a atribuição ao jornalista de tarefas de “curadoria” e de “co-produção” de conteúdo juntamente com as máquinas (cf. pp.12), surge a seguinte questão: estará efetivamente a tradicional posição do jornalista enquanto *gatekeeper* a dar lugar a um cargo de *gatewatching*?

Rafael Reis (2016) refere que “o Gatewatching passa a ter uma relevância ainda maior” (*ibid*, 2016: 83). O novo paradigma de trabalho retira do jornalista a capacidade de escolher o que noticiar e de “fazer a notícia”, num contexto em que “os consumidores estão diretamente envolvidos na produção” (Bardoel e Deuze, 2001: 94 *apud ibid*, 2016: 83). O jornalista adquire antes um papel de análise e de aconselhamento do consumo noticioso do público, evitando que os leitores se fechem em bolhas temáticas.

## **1.5 – EXEMPLOS DE REDAÇÕES QUE FUNCIONAM COM FERRAMENTAS SEMÂNTICAS**

São já alguns os órgãos de comunicação social que têm vindo a adotar este tipo de tecnologias. Iuri Lammel e Luciana Mielniczuk (2012) referem o exemplo do site da BBC Wildlife. Com mais de 30 anos de produção televisiva, o canal de televisão dispõe de uma plataforma online que funciona com uma “enciclopédia multimídia” (*ibid*, 2012: 186). Sendo um órgão de comunicação ligado à vida selvagem, cada uma das mais de mil espécies catalogadas dispõem de uma página personalizada, com vários elementos multimédia. Além destas, o site produz centenas de outras páginas com conteúdos sobre animais que partilhem as mesmas características, como o habitat, o comportamento ou a classificação biológica (*ibid*, 2012: 186). Além disso, a navegação é primeiramente orientada por um navegador central permanente, e a partir daí o visitante tem ao seu dispor um conjunto de links contextualizados com a informação que está a consultar no momento (*ibid*, 2012: 187). A integração das tecnologias semânticas neste projeto é descrita da seguinte forma:

“O funcionamento do sistema empregado na publicação e no gerenciamento do site BBC Wildlife adota várias tecnologias que caracterizam a proposta de Web Semântica, tais como o RDF e as ontologias, além da técnica de *tagging* manual, que enriquece o repositório de metadados utilizados nas operações semânticas (Lammel, 2011) (...) Além disso, outra contribuição das tecnologias semânticas foi o reaproveitamento de dados originados em sites externos, tais como descrições e estatísticas sobre espécies. Essa interoperabilidade de dados entre sites ocorre principalmente devido à padronização dos significados adotados por ambos, através dos vocabulários compartilhados.” (*ibid*, 2012: 191).

Daniela Bertocchi (2010) dá o exemplo do site do *The Wall Street Journal*, que na sua secção de crítica de cinema puxa vários termos do artigo para uma secção lateral denominada “Films Mentioned in This Article” organizando cartões informativos e organizados (*ibid*, 2010: 8). Outro exemplo é dado a partir dos sites do *The New York Times* e do espanhol *LaInformación*, onde os leitores, a partir do texto, podem seleccionar palavras específicas para compreender o seu significado, através da aplicação semântica Head Up (*ibid*, 2010: 9).

## II - METODOLOGIA

A metodologia desta investigação passa em grande medida por entrevistas, sendo este o método selecionado para a recolha de dados primários. Por um lado, foram entrevistados investigadores na área da Engenharia Informática com conhecimentos em Web Semântica, de forma a elucidar sobre aquilo em que a Web Semântica consiste, bem como os seus riscos e potencialidades. Este grupo corresponde a uma das várias categorias de interlocutores que, de acordo com Quivy e Campenhoudt (1998), devem ser entrevistados: “docentes, investigadores especializados e peritos no domínio de investigação implicado pela pergunta de partida” (*ibid*, 1998: 71). Além disso, estas entrevistas, a dado momento, adquiriram um tom exploratório. Para os mesmos autores, as entrevistas de carácter exploratório vêm no sentido de “revelar determinados aspetos do fenómeno estudado em que o investigador não teria espontaneamente pensado por si mesmo e, assim, completar as pistas de trabalho sugeridas pelas suas leituras” (*ibid*, 1998: 69), e as entrevistas a investigadores também foram formuladas para sustentar (quer corroborando quer refutando) os autores lidos.

Por outro lado, as entrevistas recaíram também sobre jornalistas, editores e diretores de jornais com presença online e com conhecimentos em Web Semântica, de forma a retirar uma perspetiva diferente destas tecnologias face ao primeiro grupo de entrevistados, bem como a perceber como é que estas tecnologias estão a ser aproveitadas nas redações ou até mesmo se estão de alguma forma a funcionar em tais espaços. Estas entrevistas serviram ainda os objetivos de aferir conhecimentos sobre a Web Semântica junto dos quadros dos jornais, recolher predisposições para conhecer mais sobre a Web Semântica, perceber a recetividade à integração destas tecnologias no setor, e fornecer um panorama geral da imprensa online em Portugal (em concordância com o título da presente dissertação). Apenas um entrevistado deste grupo afirmou não ter conhecimentos acerca da Web Semântica, mesmo assim foi incluída neste estudo para perceber até que ponto é que na sua publicação o entrevistado estaria disposto a saber mais sobre estas tecnologias e partilhar os seus conhecimentos com os demais colegas de trabalho.

Em ambos os grupos de entrevistados procedeu-se a entrevistas semi-diretivas. Foram conversas orientadas por “uma série de perguntas-guias, relativamente abertas”<sup>5</sup> (*ibid*, 1998: 192). A escolha da entrevista está relacionada com o facto de ser o método que, de forma mais explícita e aprofundada, dá a conhecer os pontos de vista de uma determinada população sobre determinados fenómenos ou acontecimentos, e na presente dissertação pretende-se precisamente retirar os pontos de vista dos entrevistados no que toca à Web Semântica e à sua ligação com o jornalismo.

---

<sup>5</sup> Na secção de Anexos estão disponíveis dois modelos de questionários utilizados nas entrevistas (Anexos C e D). Note-se que, para cada entrevista, foram feitas perguntas de repique, de forma a adaptar o fluxo da entrevista ao entrevistado.

“Interviewing gives us access to the observations of others. (...) We can learn also, through interviewing, about people’s interior experiences. We can learn what people perceived and how they interpreted their perceptions. (...) Interviewing gives us a window on the past.” (Weiss, 1994: 10).

A escolha deste método de investigação prende-se também pelo facto de permitir um maior “grau de profundidade dos elementos de análise recolhidos”, bem como a recolha de testemunhos e de interpretações dos interlocutores, “respeitando os próprios quadros de referência” (Quivy e Campenoudt, 1998: 194).

No entanto, o método por entrevista não vale por si só, pelo que “está sempre associado a um método de análise de conteúdo” (*ibid.*, 1998: 195). Trata-se aqui de um método que “oferece a possibilidade de tratar de uma forma metódica informações e testemunhos que apresentam um certo grau de profundidade e de complexidade” (*ibid.*).

As entrevistas explorarão primeiramente que conhecimentos têm os entrevistados acerca da Web Semântica, seguindo-se posteriormente questões sobre as possíveis mudanças nas suas práticas e rotinas ao longo dos últimos anos e se, de alguma forma, o trabalho tem vindo a ter em conta as recentes alterações na Web. No fundo, e entrando em concordância com o título da presente dissertação, esta fase da investigação pretende, de forma mais alargada, perceber o panorama jornalístico nacional face às tecnologias semânticas.

A escolha de métodos qualitativos de investigação prende-se com o facto de permitirem extrair informação com maior rigor, detalhe e profundidade que numa metodologia quantitativa. Além disso, a própria investigação e a reduzida população em estudo permitem uma recolha e análise de dados mais pormenorizada. O mesmo não poderia ocorrer, recorrendo a métodos de investigação quantitativa.

“Quantitative studies pay a price for their standardized precision. Because they ask the same questions in the same order of every respondent, they do not obtain full reports. Instead, the information they obtain from any one person is fragmentary, made up of bits and pieces of attitudes and observations and appraisals.” (Weiss, 1994: 11).

Contudo, a dissertação também se apoia em dados secundários quantitativos. Para suportar alguns autores foi necessário recorrer ao Digital News Report 2018 do Reuters Institute for the Study of Journalism e da Universidade de Oxford. Trata-se de um estudo desenvolvido anualmente no âmbito do Digital News Project, que desde 2013 conta com o apoio da Google. Para este projeto contribuem investigadores, patrocinadores e colaboradores de vários países, com o objetivo de perceber como se comporta o consumo de notícias num conjunto de países, do qual Portugal faz parte desde 2015. O processo de obtenção dos dados para este relatório passou por um questionário online, lançado entre o final de janeiro de 2018 e inícios de fevereiro do mesmo ano. Para alguns temas específicos,

nomeadamente o pagamento para consumo de notícias e a publicidade digital, foram também conduzidos focus groups nos Estados Unidos, no Reino Unido, na Finlândia e em Espanha<sup>6</sup>.

Quivy e Campenoudt destacam que método de investigação por inquérito permite “o conhecimento de uma população enquanto tal: as suas condições e modos de vida, os seus comportamentos, os seus valores e as suas opiniões” (*ibid*, 1998: 189). Neste caso, pretende-se conhecer os hábitos de consumo de notícias de uma população internacional. Um outro aspeto levantado pelos autores tem a ver com o facto de o método por inquérito permitir “a análise de um fenómeno social que se julga poder apreender melhor a partir de informações relativas aos indivíduos da população” (*ibid*, 1998: 189). No caso deste estudo, e da forma como ele servirá esta investigação, serão apresentados os dados relativos à influência dos algoritmos no consumo de notícias.

## 2.1 – SOBRE OS ENTREVISTADOS

Vítor Basto Fernandes, 46 anos, Professor Auxiliar no Departamento de Ciências e Tecnologias da Informação (ISTA) do ISCTE-IUL. Doutorado em Informática pela Universidade do Minho, o professor é também investigador em várias áreas, sendo a Web Semântica uma delas. Vítor Fernandes trabalhou enquanto engenheiro de qualidade na Vision Inc, no Reino Unido, cofundou em Portugal a PSI-Information Systems Lds, e este à frente do Centro de Investigação em Informática e Comunicações do Instituto Politécnico de Leiria, e colaborou em vários projetos internacionais nas áreas de integração de sistemas de informação, filtragem anti-spam e otimização multiobjetiva.

Luís Miguel Botelho, 58 anos, Professor Associado no Departamento de Ciências e Tecnologias da Informação (ISTA) do ISCTE-IUL. Além de professor é também investigador no Instituto de Telecomunicações-IUL (ISTA), e a sua área de investigação é a de Agentes Inteligentes e Inteligência Artificial. Luís Botelho esteve envolvido em diversos projetos europeus (por exemplo, AgentCities e CASCOM) nos quais participou com diversos parceiros internacionais (e.g., Imperial College, EPFL (Ecole Polytechnique Federale de Lausanne), DFKI (German Research Center for Artificial Intelligence) Philips, British Telecom, Fujitsu, e Inesc-id, na definição e implementação de diversas tecnologias e aplicações para agentes inteligentes. A sua investigação em Inteligência Artificial tem por objetivo a geração automática de histórias, isto é, a conceção e implementação de programas capazes de escrever histórias, idealmente histórias que tenham interesse para quem as lê e que sejam escritas de forma correta, refletindo o estilo do programa que as escreve.

Bruno Contreiras Mateus, 40 anos, Head of Digital do jornal Dinheiro Vivo, pertencente à Global Media. Bruno esteve também presente na criação de um departamento de Inovação e Novos Negócios no mesmo

---

<sup>6</sup> A metodologia deste relatório encontra-se na página 6 do mesmo, acessível a partir do link <http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475>.

grupo de comunicação. Enquanto jornalista passou pelo Diário do Barreiro, Público, Revista Sábado, e pelo Correio da Manhã, onde desempenhou posteriormente a função de Editor Adjunto Multiplataforma. No Global Media Group começou enquanto Editor Multimédia, ficando responsável pelos sites de várias publicações do grupo.

João Pedro Pereira, 34 anos, jornalista do jornal Público desde 2007, cobrindo as áreas de Tecnologia, Inovação e de Empreendedorismo. João também passou pelo Jornal da Universidade de Coimbra, onde ocupou o cargo de Editor-Geral, e foi também autor de uma série de artigos para a BBC. Para além de jornalista, João ocupa também atualmente a função de Coordenador de Inovação no Público, onde teve em mãos os três projetos do jornal cofinanciados pela Google.

Diogo Queiroz de Andrade, 45 anos, antigo diretor-adjunto do jornal Público. Diogo é uma das caras por detrás da criação do Observador, em 2013, juntamente com David Dinis, antigo diretor do Público. Esteve ainda no nascimento do jornal online ECO, pouco tempo antes de assumir o cargo que ocupou até junho de 2018.

Nuno Carregueiro, 40 anos, editor executivo do Jornal de Negócios. Licenciado em Gestão pelo ISEG, Nuno entrou no jornal, então denominado de Canal de Negócios, em 1999. Na referida publicação já ocupou os cargos de editor online e sub-diretor.

Leonardo Xavier, 35 anos, membro da Direção Técnica do jornal Observador. Ao longo da sua formação passou pela Engenharia Informática e terminou com o mestrado em Design e Comunicação Visual. Leonardo cofundou o primeiro espaço de *coworking* em Lisboa, o Liberdade 229. Além disso, faz parte das empresas Quodis e Frames, fundadas pelo próprio.

Hilário Lopes, 56 anos, Diretor do Arquivo Digital da RTP. Hilário iniciou a sua carreira na RTP, enquanto engenheiro, em 1985. A partir de 2000 passou para os arquivos da estação pública, estando desde então envolvido em atividades como a indexação, catalogação, edição e publicação na Internet dos arquivos da RTP, e estando ligado ao recente lançamento do Arquivo Digital, a 6 de março de 2017, data em que a estação comemorou o seu 60.º aniversário.

José Nuno Pereira, 49 anos, estudou inicialmente engenharia sem, no entanto, terminar. Foi dirigente estudantil (AEIST e AAL), e foi também eleito para a Direção do Centro de Informática do IST (CIIST), e acabou por ser um dos impulsionadores e divulgadores da Internet, acompanhando de perto o início da Rede Internet no IST e naturalmente em Portugal. No IST criou os primeiros sites e depois já como empresário e consultor criou e manteve algumas páginas institucionais. Paralelamente e mais tarde criou uma segunda carreira com trabalho em vídeo, desde a direção de fotografia até à edição de vários estilos incluindo trabalho próprio em documentário. Ambos os percursos são marcados por conhecimentos técnicos fortes, fruto de gosto, mas também educação formal. Viveu praticamente dez anos em Amesterdão nos Países Baixos. E nos últimos anos encontra-se em Lisboa a

terminar um mestrado (ISCTE) com o trabalho final focado em formatos de vídeo no sentido mais amplo do termo.

### III – PERCEÇÕES DA WEB SEMÂNTICA

Além de perceber se as tecnologias ligadas à Web Semântica funcionam dentro das redações, é igualmente importante entender como é que as pessoas ligadas ao jornalismo, bem como investigadores que ligados à área, a percecionam.

#### 3.1 – RISCOS E POTENCIALIDADES

No que toca a riscos e potencialidades da Web Semântica, em especial numa empresa de media, as respostas dos dois investigadores caminham em sentidos diferentes, na medida em que um se focou exclusivamente nos riscos, e o outro nas potencialidades.

Vítor Basto Fernandes refere que os riscos estão associados à escala dos conteúdos e da informação que temos disponível e da qual estamos cada vez mais dependentes. Com o grande volume de informação, admite a necessidade de delegar uma parte importante da sua filtragem a máquinas:

“Ao ter que delegar essa responsabilidade em máquinas nós temos que confiar que quem lhes dá esses mecanismos automáticos não está a introduzir também uma influência na própria escolha, e naturalmente que se fica mais exposto. Desse ponto de vista é perigoso.” (Vítor Basto Fernandes, 9 de abril de 2018).

Olhando para as potencialidades, Luís Miguel Botelho aponta para soluções práticas objetivas, como é o caso da geração de texto automática para artigos menos exigentes do ponto de vista literário, como é o caso dos boletins meteorológicos:

“Podemos ter programas capazes de ir buscar dados simples e produzir notícias muito simples e sem um grande rigor linguístico, até porque são tão simples que se calhar não necessitam assim de tanto rigor linguístico, como são os relatórios meteorológicos, aos quais se dedicaram várias pessoas que se ocuparam do início da geração de língua natural, isto é, de ter programas capazes de produzir texto, com base nos dados e nos modelos matemáticos e computacionais do tempo produzem, e depois gerar o relatório em que se encaixam aqueles dados todos uns nos outros e dá origem a um texto, que pode ser muito simples; em último caso até poderia ser mesmo muito simples” (Luís Miguel Botelho, 17 de abril de 2018).

Retomando o conceito de centros de reputação e dos agentes inteligentes, o mesmo investigador aponta para a utilidade de tais programas informáticos para “ver onde é que provavelmente está a verdade quando há informações contraditórias ou pelo menos diferentes” (Luís Miguel Botelho, 17 de abril de 2018), apontando que tal função poderia estar ligada à camada de *Confiança* na pilha protocolar da Web Semântica.

Entre as várias pessoas ligadas ao jornalismo entrevistadas para esta investigação, os riscos das tecnologias são geralmente minimizados. O jornalista João Pedro Pereira e o diretor técnico Leonardo Xavier afirmam abertamente não ver quaisquer riscos nestas tecnologias. Já Bruno Contreiras Mateus oferece uma visão equilibrada. O diretor digital do Dinheiro Vivo aponta como potencialidades um maior alcance dos meios de comunicação junto das suas audiências (quer seja numa lógica comercial, quer seja de um ponto de vista de serviço público), enquanto que as redações poderão servi-las cada vez melhor:

“A Web Semântica só vai aproximar-nos dos nossos leitores, numa construção que é win-win, ou seja, em que o leitor vai ganhar com rapidez o resultado, vai chegar à informação que pretende, e que nós, através de ferramentas que podem estar por detrás da disponibilização dessa informação, vamos mais depressa servir e melhor o leitor” (Bruno Contreiras Mateus, 18 de abril de 2018).

No entanto, embora não considere um ponto negativo, mas sim algo que tem de ser ponderado, Bruno Mateus levanta a questão da imediatez, em que as novas ferramentas obrigam as redações a serem ligeiramente mais lentas na disponibilização da informação; e aqui aponta para a necessidade de otimizar as ferramentas para que estas não desacelerem os ritmos de trabalho.

Leonardo Xavier sublinha a preparação dos conteúdos para que sejam consumidos em várias plataformas, conferindo-lhes uma certa “independência”:

“Cada vez mais a web está a ir na direção da leitura dos conteúdos numa só plataforma, seja o Google, seja o Facebook, seja quem for, e aí o jogo passa ou por aceitar o jogo e preparar os conteúdos para que sejam lidos por essas plataformas, ou exclui-se, fecha-se, isola-se, e recusa-se a ter o seu conteúdo disponível noutras plataformas que não sejam as suas. Para mim, o conteúdo deve viver sozinho independentemente da plataforma, e quanto melhor o conteúdo estiver classificado e preparado para ser consumido por outras plataformas e outros dispositivos, mais alcance e sucesso o conteúdo tem, e com isso mais sucesso o jornal tem” (Leonardo Xavier, 29 de maio de 2018).

Já Diogo Queiroz de Andrade aponta para as potencialidades destas tecnologias em contexto de arquivo e posterior criação de novos conteúdos:

“As tecnologias semânticas numa redação servem alguns propósitos específicos. Se eu proceder a uma devida catalogação dos conteúdos, preferencialmente usando ferramentas de inteligência artificial para o executar automaticamente e com margens de erro cada vez menores, consigo maior eficiência e relevância na utilização desses mesmos conteúdos. Logo, melhor

a pesquisa, o acesso, o relacionamento de informação e a produção de novos conteúdos.” (Diogo Queiroz de Andrade, 14 de junho de 2018)

O antigo diretor-adjunto do Público refere que assim se atingem três objetivos: “facilitar a produção jornalística, melhorar a gestão dos conteúdos, e servir melhor o leitor” (Diogo Queiroz de Andrade, 14 de junho de 2018).

### **3.2 – CONHECIMENTOS SOBRE WEB SEMÂNTICA NAS REDAÇÕES**

Um dos objetivos junto dos quadros dos meios de comunicação online passou também por aferir os seus conhecimentos acerca da Web Semântica. O primeiro passo para uma possível integração das tecnologias semânticas passa, acima de tudo, por saber da sua existência e daquilo que consistem.

Bruno Contreiras Mateus mostra alguns conhecimentos do tema, referindo que a Web Semântica confere significado à informação, e que tal se traduz, na ótica do utilizador, na oferta de uma listagem de conteúdos mais depurada e relevante; e sublinha ainda as atuais limitações dos sistemas informáticos em compreender os conteúdos:

“Portanto aquilo que nós tentamos e entendemos com o dar significado a toda esta informação, às palavras que nós colocamos nos nossos sites que correspondem a notícias, mas que um computador não sabe decodificá-las neste momento porque correspondem a palavras, nós tentamos dar-lhes então o significado para que haja depois correspondência numa pesquisa muito mais aproximada sobre aquilo que o leitor quer; vamos aproximar o resultado da pesquisa à necessidade do leitor.” (Bruno Contreiras Mateus, 18 de abril de 2018).

Já João Pedro Pereira apresenta um entendimento um pouco diferente, afirmando que o problema da compreensão das máquinas é uma questão já resolvida:

“Na altura da vaga da Web 2.0 era muito mais complexo para o software conseguir perceber o contexto e o conteúdo das páginas web. Eu diria que em geral, hoje parece-me que esse problema está resolvido, ou seja, não é rocket science para muitas aplicações conseguir compreender a generalidade do que é uma página web hoje, nomeadamente o que significa cada elemento significa, que um parágrafo é um parágrafo, que um título é um título.” (João Pedro Pereira, 24 de abril de 2018).

Embora não tenha fornecido explicações técnicas, Diogo Queiroz de Andrade mostrou-se familiarizado com a Web Semântica, admitindo tratar-se de uma preocupação em todos os projetos pelos quais passou. A construção de *back offices* de trabalho foi o exemplo apontado pelo ex-diretor-adjunto do Público,

considerados pelo próprio como “a grande ferramenta de trabalho jornalístico” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Também Leonardo Xavier demonstrou conhecer a Web Semântica, sublinhando as possibilidades na organização da informação, mas levantando a questão de se tratar de algo “discutível”:

“...É através de mecanismos automatizados conseguir retirar e acima de tudo tratar e agrupar e remixar informação, que é uma coisa que hoje em dia é discutível, há cada vez menos, mas tipicamente só é possível quando são utilizadas técnicas de classificação de conteúdo ou convenções de classificação de conteúdos que depois permitam extrair sentido deles.” (Leonardo Xavier, 29 de maio de 2018).

### **3.3 – PREDISPOSIÇÕES PARA CONHECER E DIVULGAR A WEB SEMÂNTICA NAS REDAÇÕES**

No caso da entrevista com Nuno Carregueiro, que afirmou não ter conhecimentos acerca da Web Semântica, foram-lhe explicados os seguintes pontos, de forma a inteirá-lo minimamente no tema:

- A Web Semântica é uma extensão da Web atual que permite que as páginas Web sejam lidas e interpretadas por computadores.
- Na prática isto traduz-se numa maior e mais precisa refinação dos resultados das pesquisas dos internautas.
- No campo da produção de notícias fala-se da entrada dos algoritmos no trabalho dos jornalistas, relegando ao profissional uma atividade de curadoria, de forma a dar aos leitores conteúdo que vá para além da sua “bolha temática”.
- Uma outra proposta passa por uma maior interação entre jornalistas e indivíduos e comunidades, trazendo novamente à tona a velha discussão do “jornalismo de cidadão”. A mesma proposta dala da criação de grandes bases de dados para agregar conteúdo criado pelos utilizadores de toda a Web para efeitos de verificação da veracidade dos factos.

No remate das entrevistas, procurou-se perceber se mostrou interesse por saber mais sobre esta “nova web” e quão predisposto está para torná-la um assunto de conversa entre os seus pares para, eventualmente, acelerar a integração destas tecnologias na sua redação.

O editor executivo do *Jornal de Negócios* afirma querer saber mais sobre a Web Semântica e discuti-la juntos dos seus colegas, por considerar a tecnologia um aspeto essencial. O jornalista afirma a importância de todos os jornalistas da redação, desde os estagiários até aos editores, estarem a par das mais recentes tecnologias, e que as tecnologias são um meio de aferir possíveis erros:

“...um jornalista não pode ficar fechado no seu mundo a escrever o seu texto como acha que é melhor, e depois acha que ninguém o lê porque não querem ou às vezes ouve-se o desabafo que ‘não leem porque são parvos!’. Não se pode dizer isso. Se não leem é porque alguma coisa está mal feita, falta lá alguma coisa. E muitas vezes é necessária a ajuda da tecnologia para perceber onde é que errou, portanto acho que é sempre bom ter sempre este tema presente.” (Nuno Carregueiro, 9 de maio de 2018).

### 3.4 – COMENTÁRIOS ÀS PROPOSTAS TEÓRICAS

Independentemente dos conhecimentos dos entrevistados em matérias de Web Semântica, foram apresentadas as três propostas teóricas levantadas no Enquadramento Teórico:

- O jornalista enquanto curador (Corrêa e Bertocchi, 2012) e a conseqüente passagem do *gatekeeping* para o *gatematching* (Reis, 2016);
- O “jornalismo semântico social” (Heravi e McGuinnis, 2015);
- O jornalista enquanto “nutricionista” da “dieta mediática” dos consumidores (a partir de Pariser, 2011).

Entre os vários entrevistados, apenas João Pedro Pereira concordou com a passagem do jornalista enquanto gatekeeper para um jornalista gatematcher. O jornalista do *Público* vê esta dinâmica como “inevitável”, contudo, o papel de gatekeeper é ainda algo valorizado para uma franja dos leitores:

“Ainda há tarefas adjacentes que são relevantes e que são valorizadas por alguns leitores. Por um lado, é preciso saber que nem todos os leitores são todos iguais. Há faixas de leitores para quem uma certa lógica de gatekeeping ainda funciona, e depois, acho que quando se pensa em gatekeeping pensa-se muito em 0 e 1, preto e branco, a informação passou ou não passou, chegou ao público ou não chegou. Obviamente isso agora já não acontece porque as coisas são mais fluídas e porque as pessoas podem ainda mais facilmente comunicar entre si, mas a importância relativa das notícias por exemplo, ainda é uma tarefa que está muito nas mãos dos jornalistas e com efeitos muito concretos.” (João Pedro Pereira, 24 de abril de 2018).

No mesmo jornal, Diogo Queiroz de Andrade prontamente discorda com esta alteração de funções e garante que no *Público* os algoritmos não decidirão que temas ou notícias deverão estar no *top of mind* dos leitores, sublinhando a importância de passar a visão da publicação sobre os acontecimentos:

“Não vamos ficar passivamente a ver o que passa no portão e a deixar ou não deixar passar. Continuamos ativamente a ter um olhar e a tomar decisões, e essas decisões são muito importantes. Nesse sentido o papel do jornalista não

mudou. O que nós fazíamos hoje era o que fazíamos há 20 anos. Continuamos a olhar para a atualidade e a tomar opções. (...) É muito importante que o jornalismo subsista e é muito importante que os vários jornalismo subsistam, porque o meu gatekeeping não é o mesmo que o do DN ou do Expresso ou da TSF, da SIC ou da RTP. E o leitor tem de ter acesso a estes vários gatekeepings, para poder ter uma análise completa da atualidade e da realidade.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Partindo igualmente da publicação para a qual trabalha, Nuno Carregueiro afirma que no *Jornal de Negócios* o jornalista permanece enquanto elemento fundamental não só no processo de criação das notícias, como também no da procura de informação que a publicação considere importante para atuais e potenciais leitores:

“Não tenho essa visão de o jornalista ser só um curador de conteúdos por já lhe aparecer toda a informação. Acho que o jornalista tem ainda um papel que não é substituível, pelo menos para já, pelas máquinas e pelos algoritmos. No tratamento da informação, quer na escolha de conteúdos que oferece aos leitores, quer na própria elaboração desses conteúdos. É óbvio que a tecnologia é cada vez mais relevante. (...) Temos uma bateria de indicadores que podemos sempre utilizar na hierarquização, e também, como percebemos a procura, também podemos canalizar a oferta nesse sentido. Obviamente que isso é tudo muito importante, mas também não substitui todo o know-how editorial que os jornalistas e os editores têm de ter a perceção de que, apesar de uma notícia ter o dobro do tráfego de outra, se como editor e jornalista entender eu a primeira está a ter mais tráfego e é mais relevante, eu não vou deixar de destacá-la com mais destaque.” (Nuno Carregueiro, 9 de maio de 2018).

Ainda na linha da discórdia, Bruno Contreiras Mateus oferece uma terceira via: a atribuição da função de curadoria apenas a uma parte dos jornalistas de uma redação, que pudesse variar a partir de uma hierarquização:

“Eu diria que em primeiro lugar exige que toda a redação digital tenha um mindset do digital, que conheça as ferramentas, que conheça o propósito, que saiba servir o seu leitor da melhor maneira, isso é o primeiro ponto. O segundo tem a ver com a prática, e a prática no dia-a-dia exige que haja jornalistas com maior rapidez na produção de conteúdo naquele momento, e outros que sirvam como curadores, mas todos têm de trabalhar em simultâneo. Depois podem alternar em dias diferentes, ou podem haver hierarquias. A estrutura que é criada por cada redação só compete à sua própria redação, mas todos eles têm de ter um mindset.” (Bruno Contreiras Mateus, 18 de abril de 2018).

O mesmo entrevistado suporta esta visão com o paradigma do imediatismo que marca as redações, onde o fluxo de informação é considerável e “não existe todo o tempo de reflexão que deveria existir e que nós gostaríamos que existisse, mas têm que existir elementos que tenham essa função, e que estejam permanentemente ativos” (Bruno Contreiras Mateus, 18 de abril de 2018).

Ao explicar a segunda proposta é colocado sobre a discussão o ressurgimento do jornalismo de cidadão. E aqui as respostas dispararam em várias direções. O jornalista João Pedro Pereira apresenta a perspetiva mais extremista em direção a uma “rejeição” da prática, embora admita que faça parte do contexto atual da comunicação:

“Acho que quando se fala num jornalismo de cidadão, a conceção de jornalismo não está a ser aplicada. Obviamente faz parte do ecossistema de comunicações de hoje que as pessoas fotografem, filmem, escrevam, publiquem e divulguem a informação, agora daí isso ser jornalismo vai um passo que nunca foi dado. Houve uma altura em que as redações estavam bastante fascinadas com isso e ficou mais que demonstrado que não se trata de jornalismo, é um outro fenómeno de comunicação.” (João Pedro Pereira, 24 de abril de 2018).

Um pouco na mesma linha, Diogo Queiroz de Andrade afirma que o cidadão pode servir o jornalismo, mas nunca fará parte dos seus mecanismos intrínsecos:

“Cada cidadão poderá num determinado momento produzir uma peça jornalística, não faz dele um jornalista. Agora, utilizarmos melhor o cidadão e perceber melhor como é que o podemos servir, isso sem dúvida.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Já Bruno Contreiras Mateus refere a importância do jornalista-cidadão, que este perdura na era da Web 3.0 e que as tecnologias semânticas poderão afinar esta prática que já remonta desde a Web 2.0:

“O que acaba por acontecer aqui é que o jornalista-cidadão é uma fonte de informação. Com a Web Semântica, a Big Data permite que haja o tratamento, a possibilidade de tratamento de um manancial de informação muito maior, e isso faz com que provavelmente a informação que advém do jornalista-cidadão seja muito mais bem tratada e seja muito mais facilmente identificada e checada como verdadeira ou não. O que eu acho é que existe aqui uma série de ferramentas podem dar voz ao tal jornalista-cidadão, sendo ele uma fonte de informação que interessa aos jornalistas e que interessa provavelmente aos cidadãos” (Bruno Contreiras Mateus, 5 de junho de 2018).

Da mesma forma, a metáfora dos jornalistas enquanto “nutricionistas” da dieta mediática dos leitores gerou várias reações, mas todos em sentido de uma certa discordância com a proposta.

Diogo Queiroz de Andrade concorda parcialmente, salvaguardando que o nutricionista dá ordens, enquanto o jornalista faz um trabalho menos restritivo para quem se dirige:

“A parte do jornalista ser um nutricionista, sim, é curadoria, mas um nutricionista dá ordens. Os nossos menus de escolha são muito mais abrangentes, são super menus. (...) Digamos que a panóplia de opções é muito maior que aquela que um nutricionista dá, tanto que o nutricionista tende a fechar as opções, nós tentamos abri-las, porque tentamos abrir olhares para o mundo.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Por outro lado, Bruno Contreiras Mateus discorda desta associação, afirmando que, para o jornalista ser um “nutricionista”, teria de estar a par das necessidades de todos e de cada um dos seus leitores. Além disso, afirma, tal metáfora pressuporia a existência de um padrão daquilo que seria o desejado para o consumo dos leitores, das suas necessidades e interesses (Bruno Contreiras Mateus, 5 de junho de 2018).

Bruno aponta também para a importância do grau de literacia jornalística dos leitores:

“Quanto maior for a literacia jornalística que as pessoas têm, mais apurada será a sua pesquisa, e principalmente mais apurada é a sua interpretação dos factos e do conteúdo de que lê. A minha necessidade de informação também advém da minha literacia, da minha capacidade para, em primeiro lugar, compreender e analisar aquilo que é a informação que me chega e eu ter capacidade para interpretá-la, tem a ver também com a minha capacidade para fazer uma determinada pesquisa e encontrar o conteúdo que eu pretendo.” (Bruno Contreiras Mateus, 5 de junho de 2018).

Voltando à posição do jornalista, o Head of Digital do Dinheiro Vivo também vê inviabilidade na expressão “dieta mediática”, pois diz que, na qualidade de profissional dos media, não pode agir em função dos leitores. O que pode fazer é simplesmente filtrar a informação associada negativamente, entre o que tem e o que não tem valor notícia:

“Ser um nutricionista é dizer ‘isso faz mal’. O consumo de notícia, com valor notícia e informativo não faz mal a ninguém. Aliás, até pode alterar padrões desviantes que as pessoas possam ter só de as fazer pensar. Não devem é ser utilizados como chocantes ou sinónimo de sensacionalismo. Mas aí mais uma vez, vamos afinar aquilo que é a pesquisa que as pessoas fazem dos conteúdos que querem ver, e daquilo que lhes é oferecido.(...) Portanto ser um nutricionista implica o bem e o mal, e eu não quero ir pelo bem e o mal, partindo do princípio que estamos a falar de valor noticioso em todo este conteúdo.” (Bruno Contreiras Mateus, 5 de junho de 2018).

## IV – O CENÁRIO DO JORNALISMO ONLINE PORTUGUÊS

Embora nenhuma das publicações abaixo apresentadas utilizem tecnologias semânticas, as entrevistas efetuadas pretenderam também perceber o seu estado tecnológico atual, de forma a perceber quão perto ou quão longe estão de poder algum dia começar a adotar “novas” linguagens de programação, adicionar novas rotinas de arquivo de artigos, otimização das plataformas...

### 4.1 – O CASO DO DINHEIRO VIVO/ GRUPO GLOBAL MEDIA

Bruno Contreiras Mateus desenha uma linha temporal, começando pela Web 2.0, que com as redes sociais abriu novas oportunidades para os órgãos de comunicação social no alcance de maiores audiências e na distribuição de conteúdo noticioso. Com aquilo que acredita ser um “fechamento” progressivo destas janelas de oportunidades, o Head of Digital do Grupo Global Media afirma que os jornais têm que explorar outras vias, a caminho da Web 3.0, começando pelos motores de busca. Portanto, SEO é uma das grandes prioridades, quer no jornal Dinheiro Vivo, quer nas restantes publicações do grupo de comunicação. Neste sentido é utilizado um widget no back-office do jornal:

“Este widget dá-nos uma série de parâmetros e indicadores que o baliza e que podem dizer se ele está a fazer bem ou mal, numa linguagem de três cores (verde, amarelo e vermelho), e através do SEO vamos conseguir, pelo menos, depurar cada vez mais aquilo que é o resultado da pesquisa que as pessoas possam fazer; sempre com a verdade e não enganando depois no resultado da pesquisa aquilo que é a expectativa que o leitor tem, porque senão o prejudicado somos sempre nós.” (Bruno Contreiras Mateus, 18 de abril de 2018).

O *tagging* dos conteúdos é outro grande aspeto em consideração:

“Nós trabalhamos tags em todos os conteúdos, que no fundo não são visíveis para o leitor, mas que são agregadores que contruímos e que permitem que nós depois, utilizando alguns programas, consigamos que os algoritmos deles possam ir beber também nestas tags e que possam criar novas sugestões para o leitor” (Bruno Contreiras Mateus, 18 de abril de 2018).

Estas tags servem não só como agregadores para o leitor, como também fornecem informações a widgets que ajudam a sugerir conteúdos aproximados aos consumidores do conteúdo.

Com os olhos postos no tema dos metadados, o Grupo Global Media dispõe de uma Direção Digital e Big Data:

“...é uma direção transversal, que antigamente se chamava Direção de Inovação e Novos Negócios, e que há bem pouco tempo passou a ter a nova

designação porque nós vamos trabalhar áreas muito específicas de Big Data, vamos trabalhar a nossa metadata, a nossa metainformação, para que possamos também, neste conjunto de necessidades que temos na Web Semântica, na Internet 3.0, de potenciar todo o conteúdo que nós temos, toda a informação, tentando dar um significado a esta informação, e isso vamos trabalhar através da metadata” (Bruno Contreiras Mateus, 18 de abril de 2018).

Trata-se de um órgão interno ao grupo, não-editorial, que fornece ferramentas às várias publicações. Numa primeira fase esta direção ajuda na compreensão das ferramentas que propõe, e essas ferramentas podem entrar na vida das publicações quer a partir da experimentação quer a partir da sua implementação efetiva nos sites do grupo:

“Eles apoiam-nos explicando-nos primeiro, dando-nos todo o apoio na perceção da utilização da ferramenta, e depois na própria utilização da ferramenta, e depois nas métricas que as ferramentas possibilitam, eles elaboram relatórios personalizados, adaptados à necessidade de cada publicação, e neste caso à necessidade de cada um dos gestores digitais de cada publicação, cada um de nós sabe as métricas que quer ver, muitas vezes elas são sugeridas com a sua compreensão, outras são nossas, e entramos aqui num acordo sobre que métricas são necessárias para a execução do nosso trabalho no dia a dia.” (Bruno Contreiras Mateus, 18 de abril de 2018).

Na rotina das publicações, os jornalistas podem contar com “engineers”, ou seja, todo o conjunto de técnicos que vão desde os designers aos engenheiros informáticos, permitindo em tempo real um afinamento das estratégias digitais, e uma posterior “entrada em consciência” e com a “verdadeira perceção” na era da Web 3.0.

## **4.2 – O CASO DO JORNAL DE NEGÓCIOS**

O Grupo Cofina, do qual faz parte o *Jornal de Negócios*, dispõe de um Departamento Técnico Online, que se ocupa do desenvolvimento dos novos sites, de novas aplicações, páginas especiais, e tudo o que se relacione com *Search Engine Optimization* (SEO), explica Nuno Carregueiro. No que toca a dinâmicas de trabalho, trata-se de um departamento que trabalha num “esforço conjunto e paralelo com a redação”, onde o estatuto editorial é sempre salvaguardado:

“...há comunicação e contributo mutuo de parte a parte, ou seja, são as redações que pedem ‘precisamos disto’ e eles respondem; também têm inputs como novas tendências e tecnologias que estão a aparecer e que nos dão também dicas, obviamente sem interferir no nosso trabalho editorial.” (Nuno Carregueiro, 9 de maio de 2018).

A otimização de conteúdos é a grande preocupação atual do Negócios, e Nuno Carregueiro partilhou a sua visão, quer na otimização em si, quer nas suas repercussões diretas para o jornal:

“Podemos ter uma boa notícia e pode ser a melhor de todas dentro do tema, mas se não tiver palavras-chave corretas, se o título for mau, se não tiver referências, se não tiver os links corretos e todos esses elementos, se calhar, não vão conseguir lá chegar se não for pelo tal tráfego direto. E lá está, nós queremos ter muito tráfego direto, que é um bom sinal, mas queremos também que nos resultados dos motores de busca sejamos os primeiros.” (Nuno Carregueiro, 9 de maio de 2018).

Quanto à filtragem de conteúdos para os consumidores de notícias, o editor executivo do Jornal de Negócios sublinha a passagem para os leitores daquilo que é a visão do jornal sobre o mundo e que, portanto, não são utilizados algoritmos com essa função de filtro. Nuno Carregueiro explica que “o conteúdo é procurado de forma editorial nas fontes que consideramos credíveis, quer agências de informação nacionais e internacionais” (Nuno Carregueiro, 9 de maio de 2018). As escolhas de que informação apresentar ao leitor é feita pelos próprios jornalistas, e a tecnologia é relegada para um plano meramente instrumental, apenas recorrida em determinados processos:

“A tecnologia é vista por nós mais como um meio para chegar ao leitor e para nos ajudar a hierarquizar a informação, mas não a usamos na procura da informação nem para segmentar conteúdo para os leitores. Na escolha da informação que oferecemos aos leitores não recorremos aos meios tecnológicos.” (Nuno Carregueiro, 9 de maio de 2018).

### **4.3 – O CASO DO OBSEVADOR**

A simplicidade, quer a nível de programação quer a nível de design, tem sido a maior preocupação do Observador desde o dia um, conta Diogo Queiroz de Andrade. Em 2013, ano de nascimento do jornal online, as questões dos algoritmos e da filtragem dos conteúdos junto dos leitores não era colocada como o é hoje, prossegue (Diogo Queiroz de Andrade, 7 de maio de 2018).

Passando para a atualidade do jornal, Leonardo Xavier esclarece que o Observador tem tido cuidados semânticos ao nível da programação, nomeadamente para fins de indexação junto dos motores de pesquisa:

“Temos cuidados semânticos por causa dos search engines, que são a maior fonte de tráfego para um jornal como o Observador, e isso materializa-se no HTML, que respeita em mínimos olímpicos algumas exigências semânticas, mas algo muito elementar. Talvez se manifeste mais ainda em questões de partilhas do Twitter em que se acaba por colocar tags com significado ou as

ferramentas da Google para classificar a notícia e escrever bem se se trata de um artigo ou de uma notícia, e quem é o autor, etc, para plataformas como o Google News ou mesmo o Twitter. Mas é um caso para que outros ou internamente os nossos sistemas possam ler e analisar e reagrupar a informação. É mais para as plataformas de partilha e pesquisa poderem interpretar e classificar corretamente a informação.” (Leonardo Xavier, 29 de maio de 2018).

Dentro do Observador, Leonardo gere a equipa interna de tecnologia, que no fundo, diz, funciona como “um hub de inovação, que tem imensa autonomia e liberdade” (Leonardo Xavier, 29 de maio de 2018).

Em modo de introspeção, o entrevistado admite que a questão da classificação do conteúdo deveria ser aprofundada, através de tags mais específicas, como as de autor, de entidades e de tópicos.

#### **4.4 – O CASO DO PÚBLICO**

Quando assumiu o cargo de diretor-adjunto, Diogo Queiroz de Andrade admite que o jornal se encontrava “estagnado a nível tecnológico”. Um dos aspetos apontados é o do back-office, que tem dez ou 12 anos e que é o resultado de várias sobreposições, tornando-o mais lento. Na sua visão, esta plataforma é a grande ferramenta do trabalho jornalístico, e tem de ser rápida, não no sentido de processamento, mas no sentido de conseguir produzir um artigo em poucos passos. Neste sentido, revelou, o jornal tem vindo a estudar a criação de um novo back-office. Pretende-se que seja uma ferramenta “do it all” (permitirá, para além da criação de texto, criar gráficos, inserir vídeos, fotogalerias...) e, garante “dará respostas muito melhores e mais eficientes às necessidades da Web Semântica” (Diogo Queiroz de Andrade, 7 de maio de 2018). A nova plataforma de criação de conteúdo jornalístico assenta na ideia de que cada unidade de produção (texto, vídeo, fotogaleria, infografia...) seja um produto autónomo:

“Ou seja, se eu faço uma fotogaleria, pode viver só enquanto elemento próprio, mas essa mesa fotogaleria pode ser integrada num artigo ou num conjunto de artigos ou num conjunto de fotogalerias. Num gráfico a mesma coisa; posso incluí-lo na homepage como posso incluí-lo no back office, ou seja, uma unidade com vida própria como um átomo, mas que também se junta com outros para formar um organismo.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Outro ponto destacado é o site, que também não primava pela rapidez. Em finais de setembro de 2017, o Público deu a conhecer um novo front-office. Diogo Queiroz de Andrade admite que com este novo site houve já uma certa preocupação em termos de Web Semântica, nomeadamente com a eliminação

de tags repetidas e com erros: “Tínhamos 90.000 tags e passámos para cerca de 20.000. Portanto foi um trabalho que nunca tinha sido feito no Público.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Um pouco à semelhança do caso do Grupo Global Media, o *Público* também criou uma ponte entre a redação e a sua equipa informática:

“Quando esta direção chegou, a primeira coisa que nós fizemos foi deitar abaixo ali uma parede que estava ali ao fundo que separava a programação do resto da redação. Isto porquê? Porque nós vemos que o jornalista tem de estar integrado. Aliás, nós contratámos vários jornalistas porque tivemos vários projetos da Google que nos obrigaram a reforçar as equipas, mas tirando esses projetos pontuais nós contratámos tanto as pessoas para o lado tecnológico como para o lado jornalístico.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

Além disso, refere o antigo diretor-adjunto, o *Público* está integrado da vertente mais tecnológica do grupo SONAE, a SONAE IM, que para além de incubadora de startups, tem a sua própria startup, a Bright Pixel, que tem ajudado o jornal a ganhar vários projetos da Google:

“Olhamos para a Bright Pixel como um fornecedor de conteúdos técnicos, tecnológicos, e os projetos nascem daqui, portanto o hub criativo somos nós, mas olhamos para eles muito mais como parceiros. É uma dupla cabeça. Agora o que é fundamental é que o drive criativo esteja na redação, seja nos developers, seja nos programadores, seja nos data journalists, o que interessa é que esteja aqui dentro.” (Diogo Queiroz de Andrade, 7 de maio de 2018).

#### **4.5 – ARQUIVO DIGITAL RTP: UM CASO DE POTENCIAL APLICAÇÃO DA WEB SEMÂNTICA**

Graças às potencialidades de arquivo das tecnologias semânticas, o Arquivo Digital RTP, lançado em 2017, surge nesta investigação como uma hipótese de entrada da Web Semântica nos media portugueses. Contudo, Hilário Lopes, diretor do arquivo, esclareceu prontamente que a plataforma não dispõe de tecnologias ligadas à Web Semântica, mas que dispõe de tecnologias avançadas de gestão de metadados, que segue um modelo europeu:

“O portal do arquivo está suportado num modelo de metadados que é proprietário nosso, da RTP. Ele decorre e está muito próximo do EUscreen, que é um portal europeu de rádio e televisão que é público, do qual a RTP foi parceira com mais 28 parceiros, televisões e grandes arquivos públicos, financiados pela União Europeia, há três-quatro anos, num projeto de três anos que desenvolveu um modelo de dados próprio para conteúdos de televisão na

web, que é por si só um modelo de web semântica, que está mapeado para o modelo da Europeia.” (Hilário Lopes, 30 de maio de 2018).

Ao contrário do Arquivo Digital RTP, o EUscreen já inclui ontologias desenvolvidas e outras ligações à Web Semântica, esclarece Hilário.

Mostrando o back-office do Arquivo, e olhando para a pilha protocolar da Web Semântica, o responsável pela plataforma afirma que esta se encontra claramente ao nível do XML, com um modelo de dados que reúne um conjunto de cerca de 20 elementos que classificam os conteúdos, tais como a data de emissão, data de publicação online, contribuições (realizador, apresentador...), dados técnicos (vídeo, duração, resolução, PAL ou NTC, cor, stereo), e todo um conjunto de metadados descritivos (programa ou notícia, título, título do programa, resumo sintético (sumário), resumo alargado (opcional), personalidades, localização geográfica,...).

Trata-se de um conjunto de informações que fornecem contexto aos conteúdos, de forma a suportar o motor de pesquisa do site, seja para pesquisas simples ou avançadas. Não obstante, a falta de ontologias e outros desenvolvimentos da Web Semântica fazem com que o arquivo fique limitado ao ambiente do próprio portal RTP, inviabilizando ligações para o exterior do domínio rtp.pt, admite.

Hilário Lopes tem conhecimentos sobre a Web Semântica e reconhece que esta daria mais visibilidades aos conteúdos do arquivo:

“É possível com a WS fazer a relação entre os recursos disponíveis em toda a Web e relacioná-los entre eles e fazer com que tudo aquilo que nós temos fique mais disponível e de forma mais simples, e de certeza que potencia a sua distribuição.” (Hilário Lopes, 30 de maio de 2018).

O diretor do arquivo considera que está a poucos passos da introdução da Web Semântica na plataforma, afirmando que tal pode ocorrer através de duas formas:

“Ou passa por projetos pilotos em que por exemplo o arquivo vai, ou a RTP ensina e depois os outros vão atrás; ou então passa por uma mudança estratégia global em todo o domínio rtp.pt em que há uma mudança para passar a usar essas tecnologias” (Hilário Lopes, 30 de maio de 2018).

## CONCLUSÕES

Numa primeira análise, pode dizer-se que, de facto, a Web Semântica tem pouca ou praticamente nenhuma influência sobre a atividade da imprensa online portuguesa, contudo apresenta uma série de potencialidades e riscos associados que podem efetivamente marcar o trabalho dos jornalistas. De facto, poucos são aqueles que, dentro das redações, conhecem este tipo de tecnologias; um facto visível ao longo desta investigação, nomeadamente na procura de entrevistados.

O tema não tem merecido atenção mediática nos últimos anos, e o número de redações cuja estratégia tecnológica poderá orientar-se para a integração destas tecnologias é também ele bastante reduzido. Apesar da produção científica e da existência de ferramentas ligadas à Web Semântica já em funcionamento, os ritmos e procedimentos de trabalho nas redações online não sofreram grandes alterações para além daquelas que o crescimento da preponderância do online no consumo de notícias tem vindo a exigir cada vez mais desde o final do século passado. Confirmando assim a primeira hipótese a montante da presente investigação.

Contudo, todas as publicações aqui analisadas dedicam-se na sua estratégia digital à otimização para motores de busca (SEO). Com os crescentes constrangimentos de plataformas como as redes sociais, as publicações estão de olho em novas oportunidades para potenciar os seus negócios. Os casos do Grupo Global Media e do *Público* apresentaram uma estratégia mais próxima da Web Semântica, com o *tagging* das peças como forma de conferir ao conteúdo uma catalogação e um arquivo mais eficaz; uma tarefa que cabe ao próprio jornalista, e que lhe estende as suas valências tradicionais de escrita de peças jornalísticas.

Embora menos surpreendente, a presente dissertação conclui que há um certo desfasamento entre o plano teórico e o plano laboral no que toca a algumas conceções da profissão do jornalista neste “novo” contexto tecnológico, desta forma corroborando a hipótese nº 2. Embora alguns autores argumentem que o jornalista está a ceder o seu lugar de gatekeeper para um gatewatcher, que faz um controlo menos apertado e cada vez menos prévio do fluxo de informação que chega às redações, os entrevistados mostram-se na sua generalidade contra esta conceção, defendendo que, apesar do crescente poder dos algoritmos, a escolha das notícias a escrever e a apresentar aos leitores continua a ser da responsabilidade de agentes humanos, dos quadros dos jornais. Assim sendo, poderá concluir-se que o jornalista enquanto gatekeeper não está necessariamente a ceder lugar a um papel de gatewatcher, a procura pela informação permanece ativa. Assim é rebatida a terceira hipótese colocada no início da presente dissertação.

O jornalismo semântico social poderá ter as suas dificuldades em avançar no plano prático, pois a fiabilidade no cidadão enquanto fonte ou mesmo enquanto jornalista-cidadão são aspetos que ainda merecem o ceticismo da classe jornalística.

As entrevistas feitas ao longo da presente investigação também provaram a inviabilidade da conceção do jornalista enquanto “nutricionista” da “dieta mediática” dos leitores, colocada como quarta hipótese de investigação. Tal metáfora coloca o profissional do Jornalismo numa posição demasiado limitada. Além disso, e tomando as palavras de Bruno Mateus sobre a ideia de que o ser humano tem uma maior propensão a consumir conteúdos chocantes, violentos ou de cariz sexual, vale a pena pensar no seguinte:

“Este conteúdo que pode ser associado negativamente é informação ou não. Tem valor noticioso ou não. Se tem eu tenho de o tratar como qualquer outra notícia. Se não tem valor noticioso, aí porque é que é notícia, e aí é que as redações se devem questionar. Agora se tem valor noticioso porque é que não o vou tratar com outra notícia qualquer? Porque é que as pessoas têm de ter um direito de acesso diferente à informação criminal, por exemplo, daquela que têm ao acesso à informação política, se saúde, de informação...?” (Bruno Contreiras Mateus, 5 de junho de 2018).

Ainda dentro das redações, assiste-se de facto a uma maior integração de engenheiros informáticos e de uma maior cooperação com os jornalistas. E este é claramente um passo importante no sentido da entrada de tecnologia semânticas nos órgãos de comunicação. A generalidade dos entrevistados refere a existência de órgãos internos ligados ao desenvolvimento tecnológico dos jornais. Mais uma vez, o Grupo Global Media e o *Público* surgem como paradigmas: No primeiro caso, foi criado o órgão não-editorial que tem vindo a tratar de questões de metadados; que propõe novas ferramentas de trabalho às suas várias publicações; e cujos jornalistas podem contar sempre que necessário; já no segundo caso, a equipa de programadores viu-se sem quaisquer paredes que os separassem da restante redação. Além disso, o jornal conta já com a participação de projetos da Google e uma colaboração com a startup Bright Pixel, dentro do grupo Sonae, que funciona como um ponto de aceleração tecnológica.

É nestas duas publicações, onde a colaboração entre programadores e jornalistas é maior, que os entrevistados demonstraram mais conhecimentos sobre a Web Semântica, oferecendo a oportunidade de confirmar a hipótese de que a aposta em programadores e web developers qualificados poderá ser uma das chaves para a integração das tecnologias semânticas nas redações, provando assim a hipótese nº 5.

Por fim, uma última conclusão tem a ver com o facto do paradigma da imediatez nos órgãos de comunicação online poderem fazer frente à entrada da Web Semântica:

“O imediatismo, a notícia de última hora, produz-nos um imediatismo muito grande e pode naturalmente ser aqui um fator crítico para o desenvolvimento e aplicação destas tecnologias. Mas uma vez que isto aconteça em pleno em todas as redações, tem que se aplicar (...) O que digo em relação ao imediatismo é que se essas ferramentas nos obrigarem a dar mais um passo, obrigam-nos a serem ligeiramente mais lentos, na disponibilização da informação. Nós temos

que fazer concessões, temos que otimizar ao máximo as ferramentas para que não tornem mais lento o nosso trabalho, mas também temos que ver o que é esta perspetiva win-win.” (Bruno Contreiras Mateus, 18 de abril de 2018).

Uma forma de combater este *trade-off* entre a imediatez e a publicação de conteúdo de qualidade e corretamente otimizado pode ser através dos modelos de *liveblog* ou *livechat*, em que a informação é adicionada em tempo direto através de troços informativos, ou mesmo através das notícias em atualização; onde se começa pode se dar um título e um lead, por exemplo, coloca-se a advertência “em atualização”, e a peça vai sendo completada e melhorada com o desenvolver dos acontecimentos. Aqui a Web Semântica poderia trazer importantes benefícios, facilitando e acelerando o trabalho de contextualização das peças, quer pelo seu potencial de arquivo fazendo chegar notícias relacionadas já publicadas, quer através de outras aplicações ligadas a outras áreas da Inteligência Artificial com a geração automática de texto. Com efeito, se este tipo de técnicas é considerado para fazer frente aos desafios tecnológicos sem largar mão da rapidez da produção das peças, é porque o paradigma da imediatez que o online trouxe guerreia com as novas exigências técnicas do jornalista, confirmando a sexta a última hipótese colocada no início deste trabalho.

## **POSSÍVEIS PORTAS DE ENTRADA DA WEB SEMÂNTICA**

Após as várias entrevistas, é possível provar que existem possíveis formas de integrar a Web Semântica na cena jornalística online de Portugal.

- O *Público*, com os sucessivos projetos em parceria com a Google e a cooperação com a Bright Pixel, a par com o desenvolvimento de um novo back office, tem aqui a oportunidade de testar e pôr em prática novas funcionalidades que se aproximem da Web Semântica, com a criação de uma plataforma que suporte linguagens de programação como a RDF e posteriormente começar a trabalhar ontologias.
- Ao longo da entrevista com Bruno Contreiras Mateus foi possível compreender que no Grupo Global Media há também uma visão de cooperação com empresas tecnológicas. “...este departamento está muito atento a startups e a empresas que trabalhem widgets ou que trabalhem em software específico ou ferramentas específicas, que nos possibilitem que o nosso conteúdo tenha um maior alcance possível e que seja mais facilmente legível. Criar o tal significado para toda esta informação de forma inteligente, através de Inteligência Artificial nas nossas máquinas para que elas também possam interpretar, no fundo, aquilo que é a necessidade do leitor, embora não o conheçamos, mas vamos conhecendo, desde que adquiramos determinados parâmetros ou padrões de utilização dele, por exemplo.” (Bruno Contreiras Mateus, 18 de abril de 2018).
- Apesar de não recorrer a tecnologias semânticas, o Arquivo Digital da RTP poderá ser uma porta de entrada para tecnologias semânticas, através das diretrizes europeias.

- O Nónio, enquanto agregador de várias publicações dos mais diversos grupos de comunicação portugueses com uma cobertura total de 70 sites, poderia servir-se de tecnologias semânticas para uma maior catalogação dos conteúdos, bem como para o seu redirecionamento para os leitores potenciais. Em resposta por email, Luís Nazaré, responsável pela plataforma, explicou que o Nónio é de origem comercial, com login comum, agregação de dados, DMP/profiling e SSP, não abordando, para já conteúdos editoriais e não recorrendo a tecnologias semânticas.

## **O QUE PODERIA SER FEITO?**

Dado o ponto de situação nacional, poder-se-ia dizer que há muito por fazer até uma integração da Web Semântica em contexto de redação jornalística. Pegando na ideia de Bruno Contreiras Mateus, a definição de um mindset orientado para a inovação tecnológica e posteriormente para a Web Semântica pode ser um primeiro passo a dar. Nunca as tecnologias semânticas entrarão numa redação onde nem sequer se sabe da sua existência e onde não há predisposição para inovar os métodos de trabalho. Isto pode conseguir-se, por exemplo, com uma alteração na formação dos jornalistas. A Academia, na área do Jornalismo, deveria contemplar no seu programa curricular, a passagem de conhecimentos acerca dos mais recentes avanços tecnológicos: é importante ter aulas de ciberjornalismo e saber como otimizar a escrita para plataformas na rede, mas é igualmente importante saber que tecnologias estão ou poderão estar por detrás desse trabalho, de forma a que, em contexto de trabalho, o jornalista possa adaptar-se mais facilmente a novas vagas tecnológicas e possa, juntamente com os departamentos responsáveis, encontrar novas e melhores soluções de trabalho para a redação onde desempenha funções.

Além disso, e à semelhança do Público, poderia haver maiores apostas nas plataformas de back office. Nelas poderiam ser incluídas de forma progressiva linguagens de programação como a RDF, permitindo aos poucos uma entrada dos ritmos de trabalho na era da Web Semântica. Para tal é necessária a contratação de pessoal ligado à engenharia informática com conhecimentos nestas tecnologias e linguagens, bem como uma maior interação entre estes e jornalistas na rotina de trabalho.

## **ÚLTIMAS CONSIDERAÇÕES: ESTUDAR A WEB SEMÂNTICA EM 2018?**

Estudar o Jornalismo à luz da Web Semântica pretende, acima de tudo, dar novos horizontes ao setor, marcado por dificuldades crescentes de financiamento, precaridade laboral, e uma certa ultrapassagem de barreiras éticas e deontológicas na busca da sustentabilidade. José Nuno Pereira aponta para uma situação que poderia ser evitada graças à Web Semântica:

“...hoje em dia esta relação das coisas está muito perturbada e tem uma interferência enorme com o dinheiro. Eu pago para o meu link ser relacionado antes dos outros e mais uma vez acho que mais uma vez a questão das relações

do dinheiro e das grandes corporações com várias entidades, por exemplo estás a ver um artigo sobre uma coisa qualquer e o artigo relacionado que te aparece se calhar não é o artigo mais relacionável, mas se calhar é o mais relacionável da rede de dinheiro onde tu estás. Isso podia-se eliminar, podia não ser assim.” (José Nuno Pereira, 4 de junho de 2018).

Neste sentido, o investigador do ISCTE refere que agora é o momento de “apontar caminhos e de se experimentar coisas diferentes”. Embora salvasse que há muito a fazer antes de apostar da Web Semântica, Nuno acredita que a discussão destas tecnologias por si só já é interessante.

“Quando for criada uma infraestrutura nova, quando forem criados motores de conteúdos novos, eles podem já ter isso em conta, e as pessoas que tratam a Web Semântica podem posicionar-se de modo a que quando se comece a discutir ou já se esteja a discutir coisas concretas, isso seja integrado. Essas pessoas têm que existir, porque vai ser preciso.” (José Nuno Pereira, 4 de junho de 2018).

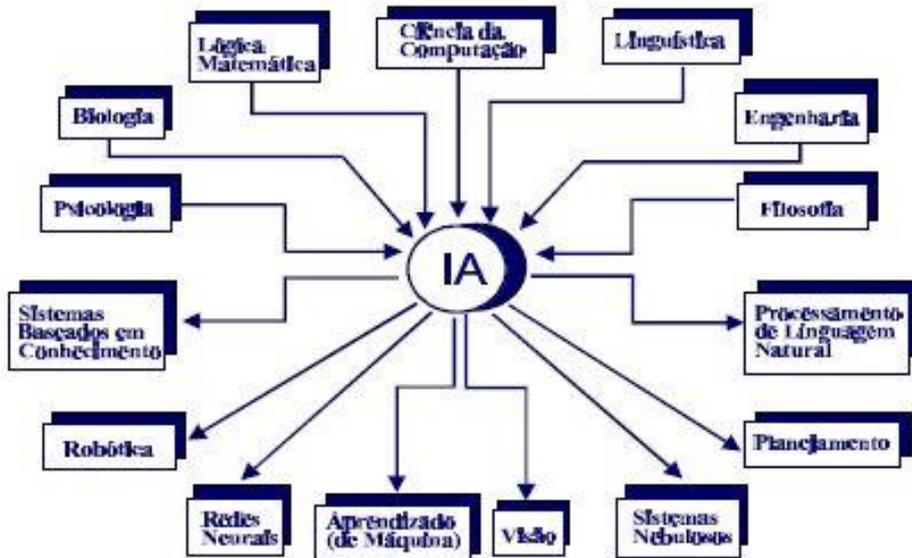
## BIBLIOGRAFIA

- Antoniou, Grigoris e van Harmelen, Frank (2008), *A Semantic Web Primer*, Londres, The MIT Press (2ª Edição)
- Berners-Lee, Tim (2000), *Weaving The Web: The Original Design and Ultimate Destiny of the World Wide Web*, Nova Iorque, HarperBusiness
- Berners-Lee, Tim *et al* (2002), “The Semantic Web”, *Scientific American Special Online Issue*, 24-30;
- Bertocchi, Daniela (2010). Narrativas jornalísticas no contexto da web semântica. Anais do II Seminário de Ciberjornalismo do Mato Grosso do Sul;
- Breed, Waren (1955). “Social Control in the Newsroom: A Functional Analysis”, *Social Forces*, vol. 33, nº 4, 326-335;
- Bruns, Axel (2011) “Gatekeeping, Gatewatching, Realimentação em Tempo Real: novos desafios para o Jornalismo”, *Brazilian Journalism Research*, 7(2), 119-140;
- Camphenoudt, Luc Van e Quivy, Raymond (2008), *Manual de Investigação em Ciências Sociais*, Lisboa, Gradiva (2ª Edição) (Edição original, 1995);
- Corrêa, Elizabeth Saad e Bertocchi, Daniela (2012), “A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria”, *MATRIZES*, nº 2;
- Flaxman, Seth *et al* (2016), “Filter Bubbles, Echo Chambers, and Online News Consumption”, *Public Opinion Quarterly* (online), vol .80, 298-320;
- Heravi, Bahareh Rahmzadeh e McGinnis, Jarred (2015), “A Framework for Social Semantic Journalism”, comunicação apresentada no 5th Annual ACM Web Science Conference, 2 a 4 de março de 2013, Paris. Disponível em <http://ceur-ws.org/Vol-1357/paper2.pdf>;
- Lammel, Iuri e Mielniczuk, Luciana (2012), “Aplicação da Web Semântica no jornalismo”, *Estudos em Jornalismo e Mídia*, nº 9;
- Lewin, Kurt (1947), “Frontiers In Group Dynamics: Channel of Group Life; Social Planning and Action Research”, *Human Relations*, vol. 1 nº 2, 143-153;
- Monard, Maria Carolina e Baranauskas, José Augusto (2000), “Aplicações de Inteligência Artificial: Uma Visão Geral”, em *Proceedings LAPTEC 2000*. São Paulo, Plêiade. Disponível em <http://conteudo.icmc.usp.br/pessoas/mcmonard/public/laptecB2000.pdf>
- Newman, Nick *et al* (2017). “Digital News Report 2017”. Disponível em [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web\\_0.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf);
- Newman, Nick *et al* (2018). “Digital News Report 2018”. Disponível em <http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475>;
- Pattal, Malik Muhammad Imran *et. al*, “Web 3.0: A real personal web! - More opportunities & more threats”, comunicação apresentada no *NGMAST 2009 - 3Rd International Conference On Next Generation Mobile Applications, Services And Technologies*, (*NGMAST 2009 - 3rd International Conference on Next Generation Mobile Applications, Services and Technologies*), 125-128. Disponível em <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5337255>;
- Pariser, Eli (2011), *The Filter Bubble: What the Internet Is Hiding from You*, Nova Iorque, The Penguin Press;
- Reis, Rafael Leitão de Paiva (2016). *O Jornalismo em Portugal e os Desafios da Web 3.0*, Dissertação de Mestrado em Ciências da Comunicação, Lisboa, Universidade Católica Portuguesa;

- Weiss, Robert S. (1994), *Learning From Strangers: The Art and Method Of Qualitative Interview Studies*, Nova Iorque, The Free Press;
- White, David Manning (1964). “The ‘Gatekeeper’: A Case Study in the Selection of News”, em Lewis Anthony Dexter e David Manning White, *People, Society and Mass Communications*, Nova Iorque, The Free Press of Glencoe;
- Wolf, Mauro (2006), *Teorias da Comunicação*, Lisboa, Editorial Presença (9ª Edição) (Edição original, 1985).  
<https://ciencia.iscte-iul.pt/authors/Vitor-manuel-basto-fernandes/cv> (02/04/2018)  
<https://ciencia.iscte-iul.pt/authors/Luis-botelho/cv> (10/04/2018)  
<https://www.brookings.edu/blog/techtank/2017/11/01/using-public-interest-algorithms-to-tackle-the-problems-created-by-social-media-algorithms/> (24/04/2018)  
[https://www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?utm\\_term=.0bf29a3d1b5a](https://www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?utm_term=.0bf29a3d1b5a) (24/04/2018)  
<http://www.digitalnewsreport.org/> (04/05/2018)  
<https://www.publico.pt/2017/03/06/culturaipilon/noticia/o-novo-arquivo-digital-da-rtp-e-uma-janela-para-a-memoria-colectiva-1764261> (01/06/2018)  
<http://leoxavier.net/about/> (01/06/2018)

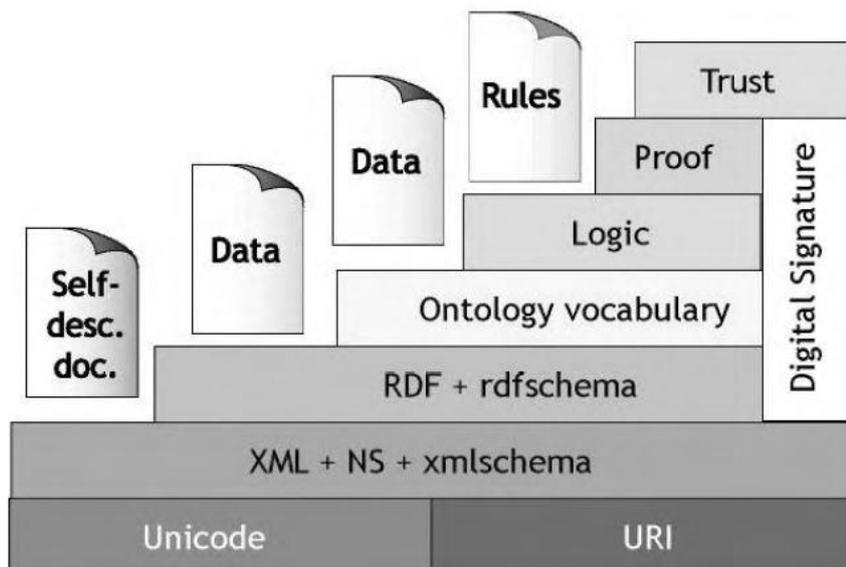
## ANEXOS

### ANEXO A – ÁREAS DA INTELIGÊNCIA ARTIFICIAL.



(Monard e Barauskas, 2000: 2)

### ANEXO B – PILHA PROTOCOLAR DE LINGUAGENS QUE SUPORTAM A WEB SEMÂNTICA.



(Antoniou e Harmelen, 2008: 19)

## **ANEXO C – MINUTA DE ENTREVISTAS PARA INVESTIGADORES.**

- Em que consiste a Web Semântica?
- Como funciona? (Esquema em pirâmide, camadas, linguagens)
- Que potencialidades e riscos podem ter as tecnologias semânticas no plano prático (para as empresas [de media])?
- O que poderia fazer um jornal para incorporar estas tecnologias?
- Estas tecnologias podem funcionar autonomamente ou necessitam/necessitarão sempre uma intervenção humana?
- Web Semântica e Web 3.0. Considera tratarem-se de duas temáticas distintas ou são praticamente o mesmo?
- Acha que a generalidade dos utilizadores da internet tem perceção do que é a web semântica e de como ela afeta as suas ações na internet? Acha que é importante que percebam?

## ANEXO D – MINUTA DE ENTREVISTAS PARA QUADROS DOS JORNAIS.

1. Sabe em que consiste a Web Semântica?
  - a. Em caso negativo, explicar com base nos seguintes pontos:
    - i. A Web Semântica é uma extensão da Web atual que permite que as páginas Web sejam lidas e interpretadas por computadores.
    - ii. Na prática isto traduz-se numa maior e mais precisa refinação dos resultados das pesquisas dos internautas.
    - iii. No campo da produção de notícias fala-se da entrada dos algoritmos no trabalho dos jornalistas, relegando ao profissional uma atividade de curadoria, de forma a dar aos leitores conteúdo que vá para além da sua “bolha temática”.
    - iv. Uma outra proposta passa por uma maior interação entre jornalistas e indivíduos e comunidades, trazendo novamente à tona a velha discussão do “jornalismo de cidadão”. A mesma proposta dala da criação de grandes bases de dados para agregar conteúdo criado pelos utilizadores de toda a Web para efeitos de verificação da veracidade dos factos.
  - b. Pedir comentários às propostas explicadas.
  - c. Em caso positivo, pedir que explique o que entende por Web Semântica
    - i. A publicação para a qual trabalha aposta em tecnologias semânticas? Se sim, como?
    - ii. (Em caso negativo) Como poderia incorporar estas tecnologias?
    - iii. Conhece casos de utilização da web semântica no jornalismo? Se sim quais? E o que pensa sobre eles.
    - iv. Que riscos e potencialidades vê nestas tecnologias semânticas dentro de uma redação?
    - v. Apresentar as duas propostas descritas em a-iii e a-iv e pedir comentários.
2. Num contexto de *information overload*, poderão os jornalistas ser considerados como “nutricionistas” na dieta mediática dos leitores?
3. É importante que adquiram essa posição?
4. Após esta entrevista, gostaria de ficar a saber mais sobre a Web Semântica, de discuti-la junto dos seus colegas e, porventura, aplicar as tecnologias inerentes no seu jornal? Porquê?
5. O jornal que aqui representa dispõe de algum departamento ou organismo que ajude à incorporação de novas tecnologias? Se sim, quais? (se necessário, dar o exemplo de um departamento de aceleração tecnológica ou de parcerias com empresas tecnológicas)

## CURRÍCULO

<b>INFORMAÇÃO PESSOAL</b>
Nome: João Vítor Gomes Patrício Morada: Rua de Baixo, nº 153, Atouguia, Ourém Email: joao_patricio95@hotmail.com Data de Nascimento: 9 de janeiro de 1995 Nacionalidade: Portuguesa
<b>FORMAÇÃO ACADÉMICA</b>
(2013-2016) FCSH – Universidade Nova de Lisboa <ul style="list-style-type: none"><li>• Licenciatura em Ciências da Comunicação<ul style="list-style-type: none"><li>○ Especialização em Jornalismo e Comunicação Estratégica</li></ul></li><li>• (2015-2016) Universidad Carlos III de Madrid (Erasmus)</li></ul> (2016-Presente) ISCTE – Instituto Universitário de Lisboa <ul style="list-style-type: none"><li>• Mestrado em Comunicação, Cultura e Tecnologias da Informação<ul style="list-style-type: none"><li>○ Especialização em Internet e Comunicação em Rede</li></ul></li></ul>
<b>EXPERIÊNCIA PROFISSIONAL</b>
(2016-2017) Estagiário no Jornal de Negócios (2017) Colaborador no ECO – Economia Online (2018 – Presente) Consultor de Marketing Digital na Súmte – Marketing Online
<b>EXPERIÊNCIA COMPLEMENTAR</b>
(2014-2017) Colaborador no Espalha-Factos <ul style="list-style-type: none"><li>• Redator na secção de Música</li></ul> (2016) Colaborador no site Máquina de Escrever (2016-Presente) Colaborador na PARQ Magazine