



# TENDÊNCIAS ATUAIS E PERSPETIVAS FUTURAS EM ORGANIZAÇÃO DO CONHECIMENTO

ATAS DO III CONGRESSO ISKO ESPANHA-PORTUGAL  
XIII CONGRESSO ISKO ESPANHA

*Universidade de Coimbra, 23 e 24 de novembro de 2017*

Com a coordenação de

---

Maria da Graça Simões, Maria Manuel Borges

TÍTULO

Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha

COORDENADORES

Maria da Graça Simões  
Maria Manuel Borges

EDIÇÃO

Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20

ISBN

978-972-8627-75-1

ACESSO

<https://purl.org/sci/atas/isko2017>

COPYRIGHT

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

OBRA PUBLICADA COM O APOIO DE



FLUC FACULDADE DE LETRAS  
UNIVERSIDADE DE COIMBRA



CEIS 20  
CENTRO DE ESTUDOS  
INTERDISCIPLINARES  
DO SÉCULO XX  
UNIVERSIDADE DE COIMBRA

**FCT**  
Fundação para a Ciência e a Tecnologia  
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PROJETO UID/HIS/00460/2013



## COERÊNCIA NA REPRESENTAÇÃO TEMÁTICA DE ARTIGOS CIENTÍFICOS NO REPOSITÓRIO DE SAÚDE PÚBLICA DA FUNDAÇÃO OSWALDO CRUZ

Fatima Cristina Lopes dos Santos<sup>1</sup>, Maria Cecília Magalhães Mollica<sup>2</sup>,  
Vânia Lisboa da Silveira Guedes<sup>3</sup>

<sup>1</sup>Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sérgio Arouca, [flopes@ensp.fiocruz.br](mailto:flopes@ensp.fiocruz.br)

<sup>2</sup>Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, [ceciliamollica@terra.com.br](mailto:ceciliamollica@terra.com.br)

<sup>3</sup>Universidade Federal do Rio de Janeiro, Programa de Mestrado Profissional em Letras, [vanialisboa@facc.ufrj.br](mailto:vanialisboa@facc.ufrj.br)

**RESUMO** Trabalho originado da dissertação de mestrado defendida no Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade Federal do Rio de Janeiro, em fevereiro de 2017. A pesquisa investiga a coerência semântica na indexação de artigos de periódicos na área de saúde pública, em língua portuguesa, indexados no Repositório de Produção Científica da Escola Nacional de Saúde Pública Sérgio Arouca (ENSP) da Fundação Oswaldo Cruz (Fiocruz), visando a contribuir com a interoperabilidade semântica entre os repositórios dessa Fundação. Para tal, foram analisados 10 artigos de periódicos científicos brasileiros, publicados em acesso aberto entre 2012-2014 e cadastrados no Repositório de Produção Científica da ENSP/Fiocruz. Os artigos foram processados nos softwares RankWord e WordSmith 6.0, aplicando-se depois as leis de Zipf e Ponto de Transição de Goffman para identificação dos termos de alta carga semântica dos textos. Os termos selecionados em cada artigo foram comparados nos 5 (cinco) indicadores utilizados na pesquisa: Região T de Goffman; Expressões WordSmith 6.0, Palavras-chave; DeCS (Descritores em Ciências da Saúde) e Título. Os resultados foram compilados utilizando-se o plano discursivo das orações, com ênfase em termos usados como figura (foregrounding) e em termos usados como fundo (backgrounding), auxiliando na verificação do grau de coerência entre os artigos. Os dados coletados revelam baixo grau de coerência semântica na maioria dos artigos estudados. Sugere-se uma investigação com amostra ampliada para confirmação dos dados obtidos, principalmente se tratando de artigos de periódicos publicados em português na área de saúde e sua importância na comunicação científica nacional.

**PALAVRAS-CHAVE** Coerência, Indexação Temática, Área de Saúde, Artigos Científicos.

**ABSTRACT** The study investigates the semantic coherence in the indexing of articles of journals in the area of public health, in the area of public health, in the area of public health. Portuguese language, indexed in the Scientific Production Repository of the National School of Public Health Sérgio Arouca (ENSP) of the Oswaldo Cruz Foundation (Fiocruz), aiming to contribute with the semantic interoperability between the repositories of this Foundation. For that, 10 papers from Brazilian scientific journals, published in open access between 2012-2014 and registered in the Scientific Production Repository of ENSP/Fiocruz, were analyzed. The articles were processed in the software RankWord and WordSmith 6.0, applying later the laws of Zipf and Goffman Transition Point to identify the terms of high semantic load of the texts. The terms selected in each article were compared in the 5 (five) indicators used in the research: Goffman's T region; Expressions WordSmith 6.0, Keywords; DeCS (Descriptors in Health Sciences) and Title. The results were compiled using the discourse plan of the sentences, with emphasis on terms used as foregrounding and in terms used as background, helping to verify the degree of coherence between the articles. The data collected

reveal a low degree of semantic coherence in most articles studied. It is suggested an investigation with an enlarged sample to confirm the data obtained, especially when dealing with articles published in Portuguese in the health area and its importance in national scientific communication.

**KEY WORDS** *Coherence, Thematic Indexing, Health Area, Scientific Articles.*

**COPYRIGHT** Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

## 1 INTRODUÇÃO

Para um alto grau de precisão na recuperação da informação na web - além da especificidade da linguagem de indexação - faz-se necessária a coerência semântica na representação da informação, visando a busca orientada em ambientes virtuais organizados para tal. Quando esses ambientes virtuais tratam de informação em saúde e do processo de comunicação científica, a recuperação da informação envolve o complexo modelo de negócio de periódicos científicos - principal canal de comunicação e divulgação entre pares.

Assim, este estudo analisa a gradação - redução ou aumento gradual e contínuo - da coerência semântica na indexação de artigos de periódicos na área de saúde pública, sistematizados no Repositório de Produção Científica da Escola Nacional de Saúde Pública Sergio Arouca (ENSP) da Fundação Oswaldo Cruz (Fiocruz). Do ponto de vista da função do periódico de canal formal de comunicação científica, a verificação do grau de coerência semântica na representatividade dos artigos de periódicos científicos poderá contribuir para a organização da informação, unificando padrões na representação temática e possibilitando a interoperabilidade semântica entre bases eletrônicas de dados.

A pesquisa se fundamenta em pressupostos teóricos e metodológicos das áreas de representação temática e da linguística documentária, no âmbito da Organização do Conhecimento na CI, bem como na Bibliometria, e tem como metas: (a) verificar comparativamente a coerência semântica na indexação de artigos científicos na área de saúde pública; (b) investigar a ocorrência de termos de indexação à luz de modelos quantitativos bibliométricos utilizados na indexação na Ciência da Informação; (c) fornecer evidências da importância do gênero discursivo artigo de periódico científico na área de saúde pública; (d) identificar os termos-figura e os termos-fundo dos artigos analisados para maior entendimento da informatividade de textos científicos; e) contribuir para o avanço do conhecimento acerca de fenômenos que norteiam a indexação da informação científica, particularmente sobre a coerência semântica na indexação da informação na área de saúde. As áreas de Ciência da Informação e Ciências da Linguagem oferecem sustentação para o desenvolvimento dessa pesquisa, na medida em que o desafio que se impõe é o de evidenciar a importância da coerência semântica na indexação de informações para a interoperabilidade de dados científicos.

O conceito de interoperabilidade não é recente na biblioteconomia. O compartilhamento de conteúdo entre bibliotecas é feito desde meados do século XX (Sayão & Marcondes, 2008). Porém, com a consolidação da internet, como via de acesso à informação e à educação, esse conceito tornou-se fundamental para a otimização da interação entre os repositórios. Andrade e Lara (2016) definem a interoperabilidade como a capacidade de dois ou mais sistemas trocar e usar informações sem um esforço especial de qualquer dos sistemas. Sendo assim, garantir a interoperabilidade torna-se o desafio de projetar serviços coerentes para uma diversidade de usuários, a partir de componentes que são

tecnicamente diferentes e gerenciados por diferentes organizações, exigindo um sofisticado grau de cooperação (Sayão & Marcondes, 2008).

A produção, a representação e a organização do conhecimento científico, especialmente publicado em artigos de periódico, gênero discursivo reconhecido como canal formal de comunicação científica, deve ser priorizada com o propósito de evitar buscas desordenadas e recuperação de informações irrelevantes. No que se refere à produção e publicação da informação científica, Asnake diz que:

Quando os resultados de um estudo de pesquisa não são publicados, outros pesquisadores não poderão apreciar o valor das evidências geradas, nem poderão construir sobre elas, e o conhecimento científico não poderá, de modo geral, crescer nem se desenvolver (Asnake, 2015, p.1972).

Para Marcondes e Campos (2008), existem dois obstáculos para o acesso e a utilização desse conhecimento, em larga escala: (i) grande número de publicações, impulsionado pela explosão informacional após a WWW; (ii) falta de estruturação desse conhecimento, tornando-o legível somente por pessoas. Berners-Lee, Hendler e Lassila (2001) mencionam que a maior parte do conteúdo disponibilizado na Web é projetada para a leitura por pessoas; eles ainda acrescentam que esse formato não estruturado impede que o conteúdo de um artigo científico seja processado por programas. Nesse sentido, os autores propõem a Web Semântica, como uma web mais estruturada, contendo informações com significados bem definidos.

No artigo *The Semantic Web*, Berners-Lee *et al.* (2001) esclarecem que os computadores deverão ter acesso a coleções estruturadas de informações e conjuntos de regras e inferências para conduzir o raciocínio automatizado, especificamente na representação e na recuperação da informação, onde o conteúdo de um texto será identificado e compreendido, semanticamente, por um programa eletrônico. Lancaster, ao reportar-se à representação da informação, menciona que:

As representações de textos consistirão no próprio texto completo, partes dele ou outra forma de representação construída por meios humanos ou automáticos. As representações de pedidos serão termos, termos apresentados em relações lógicas, enunciados textuais ou itens (Lancaster, 2004, p.284).

Na Web Semântica, um dos principais sistemas de organização do conhecimento - embasado na padronização terminológica - é a Ontologia, que possui funcionalidades que permitem que a máquina processe o raciocínio, automatizado, através de regras e inferências (Sales, Campos & Gomes, 2008). Segundo Guedes (2015), na estruturação de uma ontologia, deve-se definir um vocabulário comum, a partir do entendimento compartilhado formalmente pelos membros de uma comunidade de discurso científico. A autora esclarece que “as ontologias têm foco no estabelecimento de relações. Para tanto, são estruturadas linguagens formais que possibilitem a compreensão e o processamento, por parte das máquinas, de tais relações” (Guedes, 2015).

Como um dos pilares da Web Semântica, as ontologias são pensadas para viabilizar a interoperabilidade entre sistemas heterogêneos (conceitos) na Web semanticamente mais rica (Pickler, 2007). Marcondes e Campos (2008, p.110) afirmam que “quando o conhecimento é representado em uma linguagem declarativa, o conjunto de objetos que podem ser representados é chamado de universo do discurso”.

No que tange à indexação temática, como um dos procedimentos de representação temática, Lancaster (2004, p.10) afirma que “quanto mais especializada a clientela de um centro de informação, maior a

probabilidade de que a indexação possa e deva ser feita sob medida, ajustando-se com precisão aos interesses do grupo”. Quando uma ciência elabora seus conceitos e escolhe os termos que irão representá-los, ela constrói a terminologia da comunidade de discurso. Nesse sentido, ela se apresenta à comunidade científica indicando como gostaria de ser reconhecida e legitimada, especialmente por meio de um discurso específico, composto por termos e conceitos criados para denominar seus objetos concretos e abstratos. Benveniste diz:

Uma ciência só consegue se impor na medida em que define seus conceitos, por meio de denominações. Não existe outro modo de estabelecer sua legitimidade sem especificar, denominar seu objeto (Benveniste, 2006, p.249).

Questões relacionadas ao campo da indexação temática e recuperação da informação são objetos de estudo da Biblioteconomia e da Ciência da Informação. No capítulo sobre tradução técnico-científica, redação científica e gestão de informação, escrito por Krieger e Finatto, elas mencionam que:

Documentalistas, bibliotecários, enfim, profissionais que gestionam informação técnica e científica e a colocam à disposição de usuários, sabem o quanto é importante a familiarização com determinados assuntos e terminologias. As terminologias, para esses profissionais, valem, como elementos de indexação, como expressões de busca ou como palavras-chave de linguagem documentária. Na sua apreensão e relação com o texto especializado, diferente do que se passa com o tradutor, acentuam-se aspectos conceituais e temáticos (Krieger & Finatto, 2004, p.177)

Portanto, aspectos ligados ao mapeamento de relações lógica e ontológica, conceituais e temáticas devem ser considerados nas atividades de análise documentária e de indexação, onde se busca a adequação do termo utilizado no texto à linguagem de indexação estabelecida pelo sistema. Essas atividades estão relacionadas com a linguagem natural e a linguagem documentária, onde, a partir da análise do texto em linguagem natural, os conceitos mais relevantes são traduzidos para uma linguagem de especialidade própria da área em que será representado (Sales, 2007). Andrade, ao discorrer sobre a linguagem de especialidade, diz que ela é

Caracterizada pelo seu uso por profissionais nas diversas áreas de pesquisa e atuação, representando as áreas do conhecimento, as atividades, os processos, produtos e serviços de uma determinada área, servindo de instrumento de comunicação entre os sujeitos envolvidos nesses contextos (Andrade, 2010, p.31).

A indexação temática busca identificar os assuntos contidos no texto de um documento, traduzindo-os para uma linguagem de indexação natural - palavras que ocorrem no texto - ou linguagem de indexação controlada - termos/descriptores autorizados em vocabulário controlado (Lancaster, 2004). Quanto mais específica a representação das informações identificadas no texto em análise, maior probabilidade de recuperação de informações relevantes pelo sistema.

Na indexação, termos e descritores podem ser vistos como elementos de representação e disseminação do conhecimento, mas distinguem-se pela sua natureza e características básicas. O termo é uma unidade léxica que assume valor semântico próprio numa área de conhecimento, pela razão de integrar uma comunicação especializada. O descritor é um componente de uma linguagem constituída por gestores de informação, sendo assim, artificial (Krieger & Finatto, 2004). Por isso, a etapa de tradução de um documento, que segundo Lancaster (2004) envolve a conversão da análise conceitual num conjunto de termos de indexação, é de suma importância para a representação e está relacionada com a escolha de palavras-chave e de termos autorizados em um vocabulário controlado.

O vocabulário controlado utilizado pelo Repositório de Produção Científica da ENSP é denominado Descritores em Ciências da Saúde (DeCS), produzido pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme). O DeCS foi desenvolvido a partir do Medical Subject Headings (MeSH) da United States National Library of Medicine (NLM) e é um vocabulário trilingue que serve como linguagem única de indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos e outros tipos de materiais (DECS, 2014). O uso desse vocabulário no Repositório da ENSP objetiva auxiliar o processo de indexação de textos, indexados na base de dados do repositório, contribuindo assim para a interoperabilidade entre repositórios da área de saúde.

Sales (2007) afirma que as linguagens artificiais auxiliam as atividades de representação de conteúdos informacionais, viabilizando um armazenamento de informações que poderão ser facilmente recuperáveis. Bocatto e Torquetti (2012, p. 79) defendem que “o uso de uma linguagem de indexação controlada permite a representação na indexação e a pesquisa por assunto com maior exatidão e equidade com as necessidades informacionais dos usuários”.

Desta forma, são investigados o grau de coerência semântica entre (i) o conteúdo dos artigos selecionados, no repositório da ENSP; (ii) as palavras-chave atribuídas pelos autores dos artigos em análise; (iii) os descritores escolhidos no DeCS para a indexação desse conteúdo. Lancaster (2004, p.68) acrescenta que a coerência na indexação refere-se à extensão com que há concordância quanto aos termos a serem usados para indexar o documento. O autor afirma que quanto mais termos retirados do próprio conteúdo do texto, para indexação do documento, maior será a probabilidade de uma representação de qualidade.

## 1.1 CONTRIBUIÇÃO DA LINGÜÍSTICA DOCUMENTÁRIA

A Linguística Documentária surgiu na Espanha na década de 1990, e objetiva dar subsídios aos processos de representação para organização e recuperação de informações (Andrade, 2010). O termo Linguística Documentária foi utilizado por García Gutierrez em 1990, sugerindo uma disciplina que parte do princípio de que os problemas relacionados à informação são problemas de linguagem e, sendo assim, busca seus fundamentos nas ciências da linguagem (Lara & Tálamo, 2017).

Na Linguística Documentária, a partir da análise do mecanismo de funcionamento da língua e da Linguagem Natural, são estabelecidos os elementos para construção e uso de linguagens que darão conta de organizar, representar e comunicar informações em ambientes institucionais (Andrade, 2010). Para Tálamo e Lara (2006), a Linguística Documentária impõe-se como campo de estudos dos meios de representação da informação, em cujo vértice encontram-se as linguagens de processamento e produção da informação.

Segundo García Gutiérrez (1998 apud Lara & Tálamo, 2017) a Linguística Documentária recorre à linguística, à semântica, à gramática aplicada, à gestão da informação, como aos campos afins, como a análise do discurso, a análise do conteúdo e, de modo geral, as ciências cognitivas.

Sales (2007) entende a Linguagem Documentária como um sistema de signos que visa à uniformização do uso da linguagem de especialidade, proporcionando a representação padronizada do conteúdo informacional de textos científicos, bem como a recuperação da informação pertinente. Andrade (2010, p.33) cita que as “Linguagens documentárias delimitam os significados dos seus descritores para ofertar sentidos, de acordo com as necessidades de disponibilização e uso das informações”. Para Krieger e

Finatto (2004), o desafio está na apreensão e na divulgação da informação especializada, no sentido de que os conteúdos dos documentos devem ser identificados e adequadamente registrados, por meio do processo de indexação.

Essas Linguagens são constituídas por termos que, segundo Lara (2004), são designações que correspondem a um conceito em uma linguagem de especialidade, o termo é um signo linguístico que difere da palavra, unidade da língua geral, por ser qualificado no interior de um discurso de especialidade.

Nesta pesquisa, essas questões foram analisadas sob a perspectiva da Linguística Documentária em aproximação com perspectivas teóricas e metodológicas da Sociolinguística e do Funcionalismo, na Linguística. O Funcionalismo estuda a língua como um organismo não autônomo, produto e instrumento de comunicação, de persuasão, expressão e simulação, das manifestações humanas (Oliveira & Cezario, 2007). Para Fóris (2013), a função básica da linguagem é codificar, armazenar e transferir o conhecimento acumulado pela sociedade. A mesma autora acrescenta que a estrutura da linguagem e seus complexos processos de uso são abordados de formas diferentes, por vários ramos da Linguística. Sob esse ponto de vista, independentemente do processo de indexação temática realizado em linguagem natural ou controlada, o autor, ao elaborar um texto, seleciona, da terminologia de seu campo de conhecimento, termos/conceitos que ele julga adequados para explicitação e comunicação de suas ideias e reflexões.

Oliveira e Cezario (2007), sob a concepção do Funcionalismo, afirmam que a forma linguística tende a ser motivada pela função. Assim, eles explicam que, na produção de uma situação discursiva, fazemos escolhas lexicais e estruturais de acordo com nossos objetivos para obter sucesso na comunicação e, de alguma forma, atuar sobre os interlocutores. O autor de um texto molda o seu discurso de acordo com seus propósitos comunicativos e, nesse processo, ele elabora seu texto (aqui representado pelos artigos científicos), selecionando palavras e as representando tematicamente por palavras-chave, com o objetivo de chamar a atenção do leitor para os conceitos que ele considera mais importantes no texto.

Segundo Fóris (2013, p.423), o texto, como uma rede semântica construída por conceitos que precisam estabelecer coerência entre eles, não só “codifica a informação, mas garante que ela seja distribuída através do espaço e do tempo”. Lancaster (2004, p.68) esclarece que a coerência refere-se à extensão com que há concordância quanto aos termos a serem usados para representar conceitos, no processo de indexação temática do documento. Para Moreiro Gonzalez (1993), a coerência está relacionada ao significado alcançado através da disponibilidade e da relevância das relações, na estrutura do texto, e “é necessária para a distribuição da informação no texto” (Moreiro Gonzalez, 1993, p.134).

Em suma, no texto, o autor discursa chamando a atenção para partes ou conceitos onde acredita estar o âmago da sua produção. Neste processo, faz-se um paralelo com o estudo de Hopper (1979), citado por Oliveira e Cezario, sobre orações figura e orações-fundo:

[...] ao contar uma história, os usuários constantemente embalam as informações de acordo com suas percepções acerca do fato, das necessidades do ouvinte e dos objetivos comunicativos. Os pontos principais de uma história, suas ações, são vistos como figura (foregrounding) e os comentários, avaliações ou ações secundárias são o fundo (backgrounding) (Oliveira & Cezario, 2007, p.203).

Oliveira e Cezario (2007) explicam que, enquanto as orações-figura (foregrounding) se referem a situações objetivas, dinâmicas, as orações-fundo (backgrounding) muitas vezes se referem a elementos

subjativos, estáticos. Segundo eles, isso não significa que as orações-figura (foregrounding) sejam mais importantes que as orações-fundo (backgrounding):

Podemos contar uma história para avaliar as atitudes de um participante, ou argumentar a favor de uma tese, e, assim, as orações fundo tornam-se mais importantes para atingir os objetivos comunicativos (Oliveira & Cezario, 2007, p.204).

Neste estudo, com base na perspectiva linguística de orações-figura e orações-fundo, foram analisados termos ou descritores. Acredita-se que a análise de termos-fundo e termos-figura pode ser processada e validada com o uso das Leis de Zipf e do Ponto de Transição de Goffman, associados à utilização do programa WordSmith 6.0, que possibilita a investigação qualitativa dos termos identificados, descrevendo a sua contextualização nos textos que compõem a amostra de pesquisa.

## 2 METODOLOGIA

Esta pesquisa quanti-qualitativa teve como propósito estimar o grau de coerência semântica de conceitos e termos abordados em 10 artigos científicos indexados no Repositório de Produção Científica da ENSP/Fiocruz, e está estruturada em 7 principais etapas: (i) composição dos corpora; (ii) contagem da frequência de uso das palavras que compõem cada corpus; (iii) cálculo da frequência que corresponde ao Ponto de Transição de Goffman, para cada corpus processado; (iv) delimitação da Região de Transição de Goffman, para cada corpus processado; (v) identificação das palavras com alto conteúdo temático, de cada corpus; (vi) contextualização dessas palavras, utilizando o software WordSmith; (vii) investigação da categorização discursiva dos termos, sob a perspectiva de termos-figura ou termos-fundo. Em suma, a pesquisa baseou-se na aplicação e análise de três principais conjuntos de procedimentos, representados na figura a seguir:

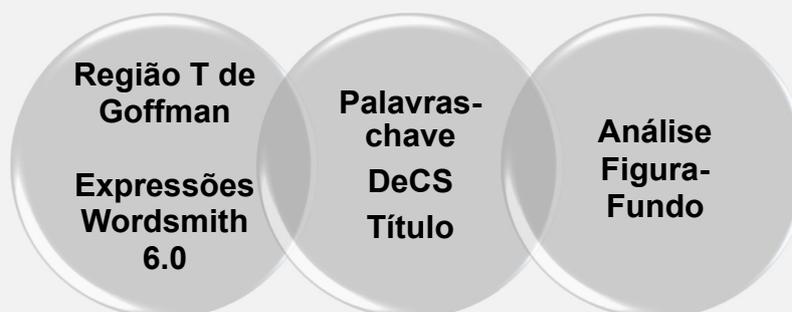


Figura 1 - Conjunto de informações estudadas nos artigos selecionados

Para a composição da amostra, foram selecionados 10 artigos de periódicos científicos sobre saúde pública, corroborando a abordagem defendida em Mollica e Guedes (2015), referente à importância do gênero discursivo artigos de periódicos, especialmente em estudos na área da Ciência da Informação. A seleção dos artigos seguiu os subsequentes critérios: (i) artigos científicos cadastrados no Repositório de Produção Científica da ENSP; (ii) artigos científicos publicados no período de 2012-2014; (iii) artigos científicos publicados em periódicos científicos nacionais; (iv) artigos científicos publicados em língua portuguesa; (v) artigos científicos publicados em periódicos científicos de acesso aberto; (vi) artigos científicos publicados em periódicos classificados entre os estratos A1 a B3 no Sistema WebQualis da CAPES.

A amostra passou pelo seguinte procedimento: verificação da frequência de uso das palavras que ocorrem nos artigos selecionados com a utilização do programa RankWords; aplicação do procedimento de delimitação da região de transição de Goffman, onde ocorrem os termos com alta carga semântica dos artigos selecionados; contextualização dos termos de alta carga semântica, localizados na Região de Transição de Goffman, em expressões/sentenças dos artigos, localizadas pelo programa WordSmith 6.0; investigação crítica da ocorrência dos termos selecionados, nos indicadores de conteúdo (região T de Goffman e Expressões WordSmith 6.0) e de representatividade (palavras-chave, DeCS e título); categorização dos resultados das atividades citadas acima, utilizando as categorias referentes ao plano discursivo das orações-figura (foregrounding) e das orações-fundo (backgrounding) (OLIVEIRA; CEZARIO, 2007), no caso desta pesquisa, termos-figura e termos-fundo dos artigos analisados; representação figurativa do conteúdo dos artigos analisados, baseando-se nos termos-figura e termos-fundo dos textos, para auxiliar a identificação da coerência semântica desses artigos.

### 3 RESULTADOS

Os artigos selecionados foram publicados em 5 títulos de periódicos científicos. Nesses periódicos, 5 artigos foram publicados em 2014, 1 artigo publicado em 2013 e 4 artigos publicados em 2012. No que se refere à classificação dos títulos no sistema WebQualis, verificou-se que os estratos situam-se entre A2 e B3. Os artigos analisados apontam o uso de 41 palavras-chave, determinando a média de 4 palavras-chave por artigo. Dessas palavras-chave, 24 foram identificadas como descritores autorizados pelo DeCS, chegando-se à média de 2 (2,4) descritores autorizados por artigo. Assim, no processamento da amostra, a porcentagem de palavras-chave autorizadas pelo DeCS é de 59% (24) e a porcentagem de palavras-chave não autorizadas pelo DeCS é de 41% (17). O aproveitamento dos termos de maior frequência de uso para elaboração de palavras-chave foi nulo em 4 textos, e nos outros 6 artigos o percentual de aproveitamento foi baixo, chegando a 14% de aproveitamento, no máximo.

A relação das palavras-chave com o DeCS foi maior. Dos artigos estudados, 6 chegaram a 100% de aproveitamento, isto é, todas as palavras-chave desses artigos são descritores autorizados por esse vocabulário controlado. Percebeu-se a preocupação com a adequação ao DeCS, sem se dar conta de que isso pode acarretar num distanciamento da representatividade do conteúdo dos textos. Para Lancaster (2004), a etapa de tradução não deve ser influenciada pelas características do vocabulário a ser usado, pois pois se o mesmo não for representado adequadamente poderá não ser localizado corretamente numa pesquisa. O texto é uma rede de conceitos que devem estabelecer coerência, se essa rede não estiver bem construída por termos representativos desses conceitos, ela poderá se tornar inconsistente (Fóris, 2013). A inconsistência dessa rede poderá prejudicar a elaboração da interoperabilidade semântica entre bases de dados. A unidade terminológica, isto é, o termo, é simultaneamente, tanto elemento constitutivo da produção do saber quanto componente linguístico, cujas propriedades favorecem a univocidade da comunicação especializada (Krieger & Finatto, 2004).

A representação figurativa do conteúdo dos artigos, sob a perspectiva teórica e metodológica funcionalista figura-fundo (Oliveira & Cezario, 2007), ratifica este resultado. Nos artigos analisados, a minoria dos termos extraídos dos textos são termos-figura, isto é, termos mais frequentes e com maior relação com o texto, indicando um baixo grau de concordância entre o conteúdo do texto e sua representatividade.

## 4 CONCLUSÃO

Percebeu-se a necessidade de um olhar mais aprofundado sobre o grau de coerência semântica na representação de artigos publicados na área de saúde pública, pois a amostra estudada resultou em baixo grau de coerência semântica na maioria dos artigos. Para Tonello, Lunardelli e Almeida Júnior (2012), a informação deve estar organizada e representada o mais fielmente possível, sendo necessária a mediação entre o conteúdo informacional do documento e aquele que dele necessita. Tal processo pode ser concretizado por meio da elaboração de palavras-chave e descritores. Segundo Krieger e Finatto (2004) a existência e a circulação de terminologias em distintos cenários comunicativos são testemunhos de que essas cumprem, prioritariamente, a dupla função de fixar o conhecimento técnico-científico e de promover sua transferência de modo pontual. Na comparação com o vocabulário DeCS, os termos de maior frequência de uso nos textos analisados, isto é, os termos-figura, não foram representados pelo vocabulário, o que leva a supor que alguns desses vocabulários podem ser feitos em bases intuitivas, sem pesquisa. Nesse caso específico, a questão pode estar na tradução/adaptação do mesmo pois, como já foi dito, o DeCS foi desenvolvido a partir do Medical Subject Headings (MeSH).

Constata-se, portanto, a importância da coerência semântica na construção de um texto científico, por conseguinte, sua análise conceitual e tradução. É importante ressaltar que na indexação de um documento não há um conjunto correto de termos, isso depende de que tipo de clientela esse documento irá atender, mas é de suma importância a pesquisa dos itens e expressões utilizadas, tanto para a construção de um vocabulário controlado quanto para escolha dos termos que irão representar um texto científico (Lancaster, 2004). Quando o texto científico - principalmente o artigo de periódico que possui grandes restrições de formato - se constrói de forma coerente, numa construção de termos que representarão devidamente os conceitos inseridos no texto, a escolha dos termos para sua representação será mais eficaz. Guedes (2010) diz que o artigo de periódico permanece no centro de uma teia formada por gêneros textuais de publicações de processo de pesquisa e que as interrelações são bem conhecidas e reconhecidas no ambiente acadêmico. Migués *et al.* (2013) dizem que o uso das palavras-chave potencializa o acesso ao conteúdo dos documentos, para além da informação que é representada pelo título e resumo; traduz o pensamento dos autores, e mantém o contato com a realidade da prática cotidiana, acompanhando a evolução científica e tecnológica, que é refletida pelos documentos.

Sugere-se uma investigação com amostra ampliada para confirmação desses dados, principalmente se tratando de artigos de periódicos publicados em português na área de saúde e sua importância na comunicação científica nacional. Os artigos de periódicos das ciências da saúde predominam na distribuição das citações por tipo de literatura (Packer, 2011), e o Brasil está em 23º lugar no ranking internacional de produção científica de artigos, ocupando a melhor posição na América Latina (Nature, 2015).

## REFERÊNCIAS BIBLIOGRÁFICAS

- Andrade, J. (2010). *A linguística documentária e a análise de domínio na organização da informação e do conhecimento*. Dissertação (Mestrado em Ciência da Informação). Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo.
- Andrade, J. & Lara, M. L. G. (2016). Interoperability and mapping between knowledge organization systems: metathesaurus-unified medical language system of the National Library of Medicine. *Knowledge Organization*, 43(2), 107-112.
- Asnake, M. (2015). A importância da publicação científica para o desenvolvimento da saúde pública. *Ciência & Saúde Coletiva*, 20(7), 1972.
- Benveniste, É. (2006). Gênese dos termos científicos. In: \_\_\_\_\_. *Problemas de linguística geral II*. 2.ed. Campinas: Pontes.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, New York, 5, May 2001. Recuperado em 2 dezembro, 2014, de [http://www.ryerson.ca/~dgrimsha/courses/cps720\\_02/resources/Scientific%20American%20The%20Semantic%20Web.htm](http://www.ryerson.ca/~dgrimsha/courses/cps720_02/resources/Scientific%20American%20The%20Semantic%20Web.htm)
- Bocato, V. R. C. & Torquetti, M. C. (2012). Interoperabilidade entre linguagens de indexação como recurso de modelagem de repertório terminológico de coordenadorias de comunicação social em ambientes universitários: uma proposta metodológica. *Informação & Informação*, 17(3), 76-101. Recuperado em 8 agosto, 2013, de <http://www.uel.br/revistas/uel/index.php/informacao/article/view/10800/pdf>
- Descritores em Ciências da Saúde (DeCS). Recuperado em 20 setembro 2014, de <http://decs.bvs.br/P/decsweb2014.htm>
- Fóris, Á. (2013). Network theory and terminology. *Knowledge Organization*, 40(6), 422-429.
- Guedes, V. L. S. (2010). *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação* (Tese de Doutorado em Linguística). Programa de Pós-Graduação em Linguística, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Guedes, V. L. S. (2015). Organização de domínios do conhecimento: taxonomias, mapas conceituais, tesouros, ontologias, folksonomias. In: Mollica, M. C. M., Batista, H. R. & Guimarães, L. S. (Orgs.). *Cybercorpora e inovação com práticas de ensinagem*. Curitiba: CRV.
- Krieger, M. G. & Finatto, M. J. B. (2004). *Introdução à terminologia: teoria e prática*. São Paulo: Contexto.
- Lancaster, F. W. (2004). *Indexação e resumos: teoria e prática* (2a. ed. rev. atual). Brasília: Briquet de Lemos/Livros.
- Lara, M. L. G. (2004). Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. *Ciência da Informação*, 33(2), 91-96.

- Lara, M. L. G. & Tálamo, M. F. G. M. (2017). *Linguística Documentária e Terminologia: experiência didática na interface das disciplinas*. Recuperado em 19 janeiro 2017, de <http://enancib.ibict.br/index.php/enancib/vii/enancib/paper/viewFile/2426/1557>
- Marcondes, C. H. & Campos, M. L. A. (2008). Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. *Ponto de Acesso*, 2(1), 107-136. Recuperado em junho 2010, de <http://www.portalseer.ufba.br/index.php/revistaici/article/view/2669/1885>
- Migués, A., Neves, B., Silva, A. L., Trindade, Á., & Bernardes, J. A. (2013). A importância das palavras-chave dos artigos científicos da área das Ciências Farmacêuticas, depositados no Estudo Geral: estudo comparativo com os termos atribuídos na MEDLINE. *InCID: Revista de Ciência da Informação e Documentação*, 4(2), 112-125.
- Mollica, M. C. M. & Guedes, V. L. S. (2015). *Gêneros discursivos e Ciência da Informação*. Recuperado em 13 maio de 2015, de <http://www.ppgci.ufrj.br/teste.html?download=213%Ageneros-discursivos-e-ciencia-da-informacao>
- Moreiro González, J. A. (1993). *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III de Madrid.
- Nature Index tables. (2015, 18 de junho). *Nature*. Recuperado de [http://www.nature.com/nature/journal/v522/n7556\\_supp/full/522S34a.html](http://www.nature.com/nature/journal/v522/n7556_supp/full/522S34a.html)
- Oliveira, M. R. & Cezario, M. M. (2007). PCN à luz do funcionalismo linguístico. *Linguagem & Ensino*, 10(1), 87-108.
- Packer, A. L. (2011). Os periódicos brasileiros e a comunicação da pesquisa nacional. *Revista USP*, 89, 26-61.
- Pickler, M. E. V. (2007). Web Semântica: ontologias como ferramentas de representação do conhecimento. *Perspectivas em Ciência da Informação*, 12(1), 65-83.
- Sales, L. F., Campos, M. L. A. & Gomes, H. E. (2008). Ontologias de domínio: um estudo das relações conceituais. *Perspectivas em Ciência da Informação*, 13(2). Recuperado em 1 julho 2010, de <http://www.scielo.br/pdf/pci/v13n2/a06v13n2.pdf>
- Sales, R. (2007). Suportes teóricos para pensar linguagens documentárias. *Revista Digital de Biblioteconomia e Ciência da Informação*, 5(1), 96-114.
- Sayão, L. F. & Marcondes, C. H. (2008). O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. *TransInformação*, 20(2), 133-148.
- Tálamo, M. F. G. M. & Lara, M. L. (2006). O campo da linguística documentária. *TransInformação*, 18(3), 203-211.
- Tonello, I. M. S., Lunardelli, R. S. A. & Almeida Junior, O. F. (2012). Palavras-chave: possibilidades de mediação da informação. *Ponto de Acesso*, 6(2), 21-34.