



# 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

2



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



## GESTÃO TERMINOLÓGICA: SUBSÍDIOS PARA A CONSTRUÇÃO DO VOCABULÁRIO CONTROLADO DA CARNAÚBA

Jacqueline A. Souza<sup>1</sup>, Monica Marques Carvalho Gallotti<sup>2</sup>, Jailma Santos<sup>3</sup>

<sup>1</sup>Professora Assistente II da Universidade Federal do Rio Grande do Norte, 0000-0001-5677-8282, [jackebci@gmail.com](mailto:jackebci@gmail.com)

<sup>2</sup>Professora Adjunta IV da Universidade Federal do Rio Grande do Norte, 0000-0002-3044-2032, [monica\\_mcg@hotmail.com](mailto:monica_mcg@hotmail.com)

<sup>3</sup>Bibliotecária do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, 0000-0001-8569-896X, [jailma.snt@gmail.com](mailto:jailma.snt@gmail.com)

**RESUMO** O contexto atual complexo traz em si a necessidade de proposição de estratégias pontuais de organização da informação e do conhecimento em variadas áreas. Diante disso, o Nutseca - Núcleo Temático da Seca e do Semiárido situado na Universidade Federal do Rio Grande do Norte - vem ao longo de 35 anos de existência reunindo mais de 100 anos de informação e coleções de documentos sobre a temática seca e assuntos relacionados. Dentre estas coleções está a uma coleção temática sobre Carnaúba. Este trabalho, em forma de pôster, visa detalhar as ações de um projeto de extensão denominado “Vocabulário Controlado da Carnaúba” financiado pelo Banco do Nordeste do Brasil. Com vistas a organizar e representar essa coleção, foi desenvolvido um projeto cujo objetivo foi: 1) compor um vocabulário controlado da carnaúba, aplicando os princípios da Linguística de Corpus (LC) e ferramentas de gestão terminológica. Especificamente visou: 2) determinar quais premissas da LC são fundamentais para a elaboração de linguagens documentárias, 3) relatar no âmbito do ensino de OI uma prática interdisciplinar. Assim, a pesquisa justifica-se pela necessidade de desenvolver instrumentos para representação e tratamento temático da coleção e é relevante pela prática interdisciplinar que converge teorias, técnicas e metodologia da Linguística de Corpus, Biblioteconomia e Terminologia. Para execução da pesquisa, foram aplicados os pressupostos teórico-metodológicos da Linguística de Corpus no tocante a manipulação dos dados linguísticos e, conseqüentemente, a ferramenta computacional E-Termos. Pelo exposto, foram identificadas as dificuldades e vantagens encontradas com a aplicação dos pressupostos metodológicos da Linguística de Corpus, bem como um protótipo do vocabulário controlado da Carnaúba.

**PALAVRAS-CHAVE** *Organização da Informação. Vocabulário controlado. Linguística de Corpus.*

**ABSTRACT** The current complex context fosters the need to propose specific information and knowledge organization strategies in various areas. Thus, this work in form of a poster, aims to describe experiences related to an extension project of a controlled vocabulary in the subject of a *Carnaúba*. The project “Carnauba Controlled Vocabulary” received financial support from *Banco do Nordeste do Brasil* and took place at the *Núcleo Temático da Seca e do Semiárido*, a documentation center that deals with drought at the Federal University of Rio Grande do Norte. The center exists for over 35 years and has gathered more than 100 years of information and collections of documents on the subject. In order to organize and represent this collection, the project was developed and it proposed: 1) to elaborate a controlled vocabulary of *Carnaúba*, applying the principles of Corpus Linguistics (CL) and use terminology management tools. Specifically, it aimed at: 2) determining which premises of the CL are fundamental for the elaboration of documentary languages, 3) Describe an IO interdisciplinary practice. Thus, the research is justified by the need to develop instruments

aiming thematic treatment and representation of collections, relevant as an interdisciplinary practice that converges theories, techniques and methodology of Corpus Linguistics, Librarianship and Terminology. For the execution of the research, the theoretical-methodological assumptions of Corpus Linguistics were applied in relation to the manipulation of linguistic data and, consequently, the computational tool *E-Termos*. It was possible to identify the difficulties and advantages encountered with the application of the methodological assumptions of Corpus Linguistics, as well as a prototype of the *Carnaúba* controlled vocabulary.

KEY-WORDS *Information Organization. Controlled Vocabulary. Corpus Linguistics.*

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

## INTRODUÇÃO

Caracterizado pela baixa umidade e pelo pouco volume pluviométrico, o semiárido nordestino representa 18,3% do território brasileiro e no Nordeste é formado por nove estados: Maranhão, Piauí, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia e Rio Grande do Norte. Neste cenário, há uma riqueza de plantas adaptadas ao clima seco e que são economicamente exploradas para a produção de óleos como o Catolé, Marmeleiro, Oiticica; de látex como o Pinhão Maniçoba e de ceras, como a Carnaúba. (Suassuna, 2011).

Nomeadamente, a carnaúba (*Copernicia prunifera*) ou carnaíba, carnaubeira, trata-se de uma palmeira também conhecida como a árvore da vida, tem importância econômica, social e ambiental. Acerca da economia, Carvalho & Gomes (2009) ressaltam que consiste no conjunto de atividades que utilizam as folhas, o caule, o talo, a fibra, o fruto e as raízes dessa palmeira para a fabricação de inúmeros produtos artesanais e industriais, destacando a extração do pó de carnaúba, a partir das folhas, por sua importância na geração de emprego e renda complementar no campo e por ser a matéria-prima da cera, que possui inúmeras aplicações no ramo industrial, sendo um importante elemento para redução da pobreza na região nordestina, entretanto sua exploração sofre influências do clima seco.

Diante desta realidade acima mencionada e dada a importância destes recursos faz-se necessário a elaboração de estratégias em torno da organização de informações provenientes deste contexto. Tal necessidade se dá devido ao fato de que na atualidade existe uma dispersão de informação impulsionado pelo avanço da Tecnologia da Informação e Comunicação. Organizar a informação acumulada em estoques informacionais tem se tornado um desafio e ao mesmo tempo uma solução. Nesse sentido corroboramos com o pensamento de Rowley quando afirma que “organizar a informação para melhor compreendê-la e recuperá-la é imprescindível, para isso são necessários padrões de estruturação pré-definidos, visando facilitar, agilizar e otimizar a chegada da informação ao usuário final, resultando em sua satisfação ajudando-o em seus processos decisórios (Rowley, 2008).

Diante disso, o Nutseca - Núcleo Temático da Seca e do Semiárido - situado na Universidade Federal do Rio Grande do Norte, o qual contém coleções de documentos sobre a temática seca e assuntos relacionados. Assim, este trabalho, em forma de pôster, visa detalhar as ações de um projeto de extensão que recebeu financiamento do Banco do Nordeste do Brasil e visou a criação de um vocabulário controlado sobre Carnaúba. Com vistas a organizar e representar essa coleção, foi desenvolvido um projeto cujo objetivo foi: 1) compor um vocabulário controlado da carnaúba, aplicando os princípios da Linguística de Corpus (LC) e ferramentas de gestão terminológica. Especificamente visou: 2) verificar

quais premissas da LC são fundamentais para a elaboração de linguagens documentais, 3) relatar no âmbito do ensino de OI uma prática interdisciplinar.

O desenvolvimento desta pesquisa procurou conciliar as necessidades do Nutseca para indexar sua coleção e promover maior visibilidade e acesso ao conteúdo informacional sobre a seca e o semiárido que, dado seu impacto e importância social e econômica, proporcionalmente pode-se considerar que ainda academicamente não tem o seu potencial explorado, desde pouca produção na literatura científica como sistematização para acesso a fontes de informação. No âmbito do Departamento de Ciência da Informação da UFRN, tocante ao ensino na área de Organização da Informação, também proporcionou desenvolver uma atividade prática, evidenciando a interdisciplinaridade.

## NÚCLEO TEMÁTICO DA SECA E DO SEMIÁRIDO

O Núcleo Temático da Seca e do Semiárido é um órgão suplementar ligado ao Centro de Ciências Sociais Aplicadas (CCSA) da Universidade Federal do Rio Grande do Norte (UFRN). É caracterizado como um centro de documentação especializado em Seca e assuntos relacionados. Seus principais objetivos são os de: propor estratégias em torno da captura, organização, tratamento e difusão da informação sobre a temática seca e semiárido bem como apoiar os programas de ensino, pesquisa e extensão em suas necessidades de acesso à informação especializada em seca e semiárido e dinamizar a produção científica concernente a seca e semiárido. (CARVALHO, 1998).

Seu acervo é constituído livros, periódicos, fotografias, artefatos tridimensionais, literatura cinzenta, anuários, índices, mapas, entre outros documentos compondo uma rica “ecologia” documental e massa crítica na área da seca. Seu acervo é tido como único e um dos maiores do Brasil. Este, está dividido em coleções específicas tais como: A Universidade e a Questão Nordestina, Seca e Semiárido, Carnaúba e a do Vale do Assú ou Projeto Baixo-Assú, construído numa visão epistemológica interdisciplinar. Esta vasta documentação é organizada com vistas a difusão da informação por meio de portais de informação e websites institucionais.<sup>1</sup>

## VOCABULÁRIO CONTROLADO E A LINGUÍSTICA DE CORPUS

Enriquecendo as pesquisas sobre a construção de ontologias e vocabulários controlados, Silva, Almeida & Souza (2013) consideraram que os procedimentos mais adequados são os sugeridos pelo manual disponível no sítio da Biblioteconomia, Informação e Tecnologia da Informação (BITI), autoria de Campos, Gomes & Motta, (2004):

- 1) Planejamento
- 2) Composição do vocabulário
- 3) Organização dos conceitos
- 4) Apresentação final
- 5) Avaliação

---

<sup>1</sup> Ver <https://ccsa.ufrn.br/nutseca/NUT-Seca/Nut%20Seca.html> e [http://www.siabiclientes.com/nutseca/Telas/w\\_busca\\_rapida.php](http://www.siabiclientes.com/nutseca/Telas/w_busca_rapida.php)

Com base nestas etapas para construção, para atender aos objetivos propostos neste trabalho, discorrer-se-á sobre Linguística de Corpus. Neste sentido, convém apresentar o seu conceito, assim como a definição de corpus e suas características.

A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador. (Berber Sardinha, 2000, p. 325).

Condicionada a tecnologia, a LC circunscreve-se a uma abordagem probabilística da língua, cujo armazenamento, acesso, manuseio e exploração ocorre via computador. Suas aplicações versam acerca da compilação dos corpora<sup>2</sup>, desenvolvimento de ferramentas computacionais para análise linguística, concordanciadores, para descrição de linguagem, identificação de padrões, é útil para ensino-aprendizagem de línguas, processamento de língua natural, confecção de dicionários terminológicos e lexicográficos, para análises morfossintáticas, discursivas, semânticas, gêneros discursivos e traduções.

Neste sentido, Oliveira (2009) ressalta que a LC representa uma nova abordagem filosófica para os estudos da linguagem e que não pode ser definida apenas como uma metodologia emergente, mas uma nova maneira de desenvolver pesquisas. Como uma ciência empírica fornece dados sobre a linguagem em uso para descrevê-la de forma adequada, que não se restringe apenas a compilação e coleta de dados, mas para geração de novos conhecimentos, novas descrições das línguas, entre outras atividades que envolvam a experiência humana tal como é construída na linguagem, o que é concernente a Organização e Representação da Informação.

Central a LC, a definição de corpus que incorpora suas principais características e que expõe como pode ser útil no desenvolvimento de produtos informacionais é:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise' (Sanchez, 1995, *apud* Berber Sardinha, 2000, p. 338)

Acerca desta definição, Berber Sardinha (2000) pontuou aspectos relevantes destacados a seguir:

- **A origem:** Os dados devem ser autênticos, ou seja, o corpus deve ser composto de textos em linguagem natural. Os textos não podem já ter sido produzidos para fins de pesquisa linguística. Também, deve-se observar se o texto é produzido por falantes nativos.
- **A composição:** O conteúdo do corpus deve ser criteriosamente escolhido. Os princípios da escolha dos textos devem seguir, acima de tudo, as condições de naturalidade e autenticidade. Ou seja, o conteúdo do corpus deve ser selecionado a fim de garantir que o corpus tenha uma certa característica. Por exemplo, se o desejo é construir um corpus de português brasileiro escrito que represente a língua portuguesa, tal qual ela é escrita no Brasil, em sua totalidade, a coleta deve ser guiada por um

---

<sup>2</sup> O plural de corpus é corpora.

conjunto de critérios que garanta, entre outras coisas, que o maior número possível de tipos textuais existentes no português brasileiro esteja representado, que haja uma quantidade aceitável de cada tipo de texto e que a seleção dos textos seja aleatória, a fim de que não se contamine a coleta com variáveis indesejáveis.

- **A formatação:** Os dados do corpus devem ser legíveis por computador.
- **A representatividade:** O corpus deve ser representativo de uma língua ou variedade, ou seja, uma vez que a linguagem é de caráter probabilístico, há possibilidade de estabelecer uma relação entre traços que são mais comuns e menos comuns em determinado contexto.
- **A extensão:** o corpus deve ser vasto e, neste sentido, comporta 3 dimensões: o número de palavras, sendo que quanto maior o número de palavras, maior será a chance do corpus conter palavras de baixa frequência, as quais formam a maior parte das palavras de uma língua; número de textos, pois quanto mais texto, maior será a garantia de tipo textual, gênero e registro; e o número de gêneros ou tipos textuais, que permitirá maior abrangência do espectro genérico da língua.

Com base na definição de corpus e nos pontos relevantes a ser considerado na composição do corpus, podemos verificar que estes princípios são importantes na elaboração de linguagens documentárias. Por exemplo: uma vez que a linguagem documentária deve abarcar a maior variedade de descritores e não-descritores para que cumpra sua função.

## METODOLOGIA

Em termos metodológicos este trabalho inicia por uma pesquisa bibliográfica acerca dos temas da Organização da Informação, construção de vocabulário controlado e Linguística de Corpus. Tal abordagem foi necessária para se traçar as premissas conceituais a fim de fundamentar o trabalho em tela. Em seguida foi feito um relato de experiência em consonância com os postulados de Minayo (1994) e Connelly & Clandinin (1995) cuja concepção geral prevê que registrar e relatar experiências científicas é fundamental para o avanço da ciência. Para execução da pesquisa, foram aplicados os pressupostos teórico-metodológicos da Linguística de Corpus no tocante a manipulação dos dados linguísticos e, conseqüentemente, a ferramenta computacional e-Termos<sup>3</sup>, acrônimo de Termos eletrônicos, um ambiente computacional colaborativo web de acesso livre e gratuito dedicado a gestão terminológica.

---

<sup>3</sup> A ferramenta computacional é oriunda do Projeto e-Termos, um projeto acadêmico desenvolvido em parceria entre a Embrapa Informática Agropecuária (CNPTIA), Universidade de São Paulo (USP Campus de São Carlos, SP) e Universidade Federal de São Carlos (UFSCar), representados pelos laboratórios de pesquisa LabInfo (Laboratório de Organização e Tratamento da Informação Eletrônica), NILC (Núcleo Interinstitucional de Linguística Computacional) e o GETerm (Grupo de Estudos e Pesquisas em Terminologia), localizados nas três instituições, respectivamente. Informações disponíveis em: <https://www.etermos.cnptia.embrapa.br/index.php>

## RESULTADOS

O Vocabulário Controlado da Carnaúba pretende auxiliar a indexação das fontes digitalizadas, implicando um tratamento temático do acervo do Nutseca. Dada a configuração do projeto, em sua primeira fase, os resultados obtidos são a configuração da sua macroestrutura onde se observam o domínio temático estruturado através da hierarquização e seus desdobramentos, como também as relações lógicas que se baseiam na similaridade das características entre conceitos, formando, assim, uma relação de gênero / espécie. O vocabulário é composto por um conjunto com média de 400 termos que representam conceitos relacionados à Cultura da Carnaúba, os quais estão dispostos em uma estrutura em ordem alfabética, destacando-se as relações de equivalência.

Tocante as etapas para construção e gestão dos termos, foram implementadas todas as etapas sugeridas pela LC. A compilação do corpus deu-se com textos oriundos da web e com a própria coleção do Nutseca, com 296 textos autênticos e 6 gêneros distintos. Considerando a abordagem probabilística da LC, foram observados os aspectos quantitativos e qualitativos, com um corpus de 1.288.461 palavras e bom índice de riqueza vocabular.

Ainda sobre gestão terminológica, verificou-se que as etapas sugeridas para elaboração de vocabulário controlado, convergem em muitos pontos com as etapas implementadas no uso da ferramenta e-Termos.

## CONCLUSÕES

O Relato de experiências no terreno das práticas de Organização da Informação no geral em específico da construção de vocabulários controlados é uma mais valia pois permitem a socialização de procedimentos e metodologias adotadas agregando valor à base de conhecimento existente no assunto. Baseando-nos no que foi apresentado, concluiu-se que a Linguística de Corpus fornece aportes teórico e metodológico na gestão terminológica para a elaboração de linguagens documentais. Ressalta-se a importância de critérios para compilação do corpus e destaca-se o uso das ferramentas computacionais, sobretudo por promoverem a colaboratividade entre diferentes atores envolvidos no processo, como bibliotecários e os especialistas de domínios específicos que, por exemplo, são consultados para validar os termos que compõem os vocabulários e esclarecer dúvidas, dentre outros aspectos.

## REFERÊNCIAS BIBLIOGRÁFICAS

Barité, M. (2011). Sistemas de Organización del Conocimiento: una tipología actualizada; In: Sistemas de Organização do Conhecimento: uma tipologia atualizada. *Informação & Informação*, 16(2), 122-139. Doi: 10.5433/1981-8920.2011v16nesp.p122

Campos, M. L. D. A., Gomes, H. E., & Motta, D. F. (2006). *Elaboração de tesauro documentário: Tutorial*.

Carlan, E., & Medeiros, M. B. B. (2011). Sistemas de Organização do Conhecimento na visão da Ciência da Informação. *Revista Ibero-Americana de Ciência da Informação*, 4(2). Disponível em: <http://repositorio.unb.br/handle/10482/12867>

Carvalho, J. D., & Gomes, J. (2009). Pobreza, emprego e renda na economia da Carnaúba. *Revista Econômica do Nordeste*, 40(02), 361-378.

Carvalho, R. P. F. de. (1999). Núcleo temático da seca/UFRN: uso do acervo informacional. *Informação & Sociedade: Estudos*, 9(2).

Freire, I. M., Carvalho, L. M., Carvalho, M. M., & Aranha, T. D. Q. (2008). Ampliando o acesso livre a informação: a digitalização do acervo do Núcleo Temático da Seca. *Informação & Sociedade*. João Pessoa, v.18, n.2, p. 137-142, maio/ago. Disponível em: <http://www.repositorio.ufrn.br:8080/jspui/handle/1/3125>

Lima, J.L.O. & Alvares, L. (2012). *Organização e representação da informação e do conhecimento*. In: L. Alvares (org), *A organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações* p. 21-47. São Paulo.

Oliveira, L. P. (2009). Linguística de Corpus: teoria, interfaces e aplicações. *Matraga-Revista do Programa de Pós-Graduação em Letras da UERJ*, 16(24). doi: 10.12957/matruga

Rowley, J. E., & Hartley, R. J. (Eds.). (2008). *Organizing knowledge: an introduction to managing access to information*. Ashgate Publishing, Ltd.

Sardinha, A. P. B. (2001). Uso de corpora na formação de tradutores. *Delta*, 19, 43-70.

Sardinha, T. B. (2000). Corpus linguistics: history and problematization. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 16(2), 323-367. Disponível em: <http://bit.ly/2sx8hld>

Schiessl, M., & Shintaku, M. (2012). *Sistemas de organização do conhecimento*. In: Alvares, L. (org), *A organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações* (49-118). São Paulo.

Silva, D. L. D., Souza, R. R., & Almeida, M. B. (2013). *Uma comparação de metodologias para construção de ontologias e vocabulários controlados*. Disponível em: <http://hdl.handle.net/123456789/965>

Silva, D. L. D., Souza, R. R., & Ramos, J. A. D. A. (2010). Representação da informação e do conhecimento em sistemas de informações: contribuições de campos interdisciplinares. Disponível em: <http://hdl.handle.net/123456789/914>

Standars, N. I. S. O. (2005). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*.

Suassuna, J. (2011). Semi-árido: proposta de convivência com a seca. *Cadernos de Estudos Sociais*, 23(1-2).