



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



PROPOSTA METODOLÓGICA PARA MAPEAMENTO E ANÁLISE DE CONFIGURAÇÕES DE DISCURSOS SOCIAIS: AS REPRESENTAÇÕES DISCURSIVAS DE DILMA ROUSSEFF

Ludmila Salomão Venâncio¹, Renato Rocha Souza²

¹Universidade do Estado de Minas Gerais – UEMG, ludmila.salomao@uemg.br

²Fundação Getúlio Vargas – FGV, renato.souza@fgv.br

RESUMO Propõe-se uma abordagem metodológica para compreensão do universo semântico e dos posicionamentos dos sujeitos em discursos sociais, evidenciando as marcas argumentativas presentes em seus pronunciamentos. Para tanto, analisam-se textos proferidos pela ex-presidente da República do Brasil, Dilma Rousseff. A metodologia consiste em uma análise lexicométrica e uma análise de redes terminológicas que, viabilizadas pela identificação e pela extração automática das unidades lexicais mais frequentes em cada texto e pela construção de redes, apontaram diversas apropriações contextuais e o estabelecimento de significados contingentes diacronicamente. Desse modo, evidencia-se que a formalização explicativa, empreendida pela abordagem proposta, possibilita compreender melhor as configurações e remanejamentos das trocas discursivas em domínios dinâmicos.

PALAVRAS-CHAVE *representação e organização do conhecimento, análise lexicométrica, discursos sociais, discurso político, redes terminológicas.*

ABSTRACT We propose an approach to understand the semantic field and the subjects' positions in social discourses, evidencing the argumentative marks present in their pronouncements. To do so, we analyze texts delivered by the former president of the Republic of Brazil, Dilma Rousseff. The methodology consists of a lexicometric analysis and also an analysis of terminological networks that, enabled by the identification and automatic extraction of the most frequent lexical units in each text and the construction of networks, could lead us to the conclusions that several contextual appropriations and the establishment of diachronically contingent meanings. Thus, it is evident that the explanatory formalization, undertaken by the proposed approach, enables to better understand the configurations and re-assignments of discursive exchanges in dynamic domains.

KEYWORDS *representation and organization of knowledge, lexicometric analysis, social discourses, political discourse, terminology networks.*

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

INTRODUÇÃO

Ao longo dos últimos anos, a pesquisa da linguagem vem se tornando tema central no campo da representação e organização do conhecimento. Muitos dos estudos nesse tema incluem o uso de diversas técnicas automatizadas de tratamento de corpora textual como, por exemplo, o uso de processamento (automático) da linguagem natural (PLN). O PLN é uma subárea tanto da ciência da computação, quanto da linguística, e consiste no desenvolvimento de modelos computacionais para a consecução de tarefas que dependem de informação registrada em formato textual. Dentre estas tarefas, temos, como exemplo, a tradução e interpretação de textos, a busca de informações em documentos e modelagem de interfaces homem-máquina. De acordo com Marciniak *et al* (2009), a pesquisa em PLN está voltada, essencialmente, a três aspectos da comunicação em língua natural: som (fonologia); estrutura (morfologia e sintaxe) e significado (semântica e pragmática). Na área da ciência da informação, o PLN auxilia a processos de indexação de documentos, análise de assunto e organização de informações, utilizando, para tal, técnicas variadas, como a extração de entidades nomeadas e identificação de estruturas semânticas, como os sintagmas nominais, para a indexação e recuperação da informação. Outra das muitas ferramentas que vêm auxiliando as pesquisas em linguagem são os sistemas de organização do conhecimento (SOC/KOS). Como SOCs, consideramos todos os tipos de esquemas para organização de informação e promoção da gestão do conhecimento (Hodge, 2000) sendo os mesmos representações do conhecimento baseadas em conceitos com diferentes graus de detalhe dentre os relacionamentos entre estes (Souza et al, 2012). SOCs nos auxiliam a mapear e guiar a extração de informações em domínios específicos do conhecimento e na adoção de esquemas representacionais ontológicos em sistemas de organização e recuperação da informação. Ferramentas como os tesouros e as ontologias auxiliam processos de representação de relacionamentos conceituais e semânticos; uma proporção significativa dos estudos bibliométricos, permitindo também que frequência de termos em documentos e construção de índices sejam menos sujeitos a variações lexicais e de sinonímia. Finalmente, a união das técnicas de PLN com os SOCs permite a construção de processos de indexação automática realizada com base em termos escolhidos, dentre aqueles que se mostram relevantes em um determinado domínio.

Em todos esses casos, a linguagem torna-se o fundamento básico para a proposição de metodologias e construção de ferramentas para a representação e a organização do conhecimento, atentando-se para seu emprego pelos diferentes sujeitos sociais mobilizados em práticas discursivas diversas (Venâncio, 2014).

Neste contexto, a presente comunicação tem como objetivo apresentar uma proposta metodológica para mapeamento e análise do universo semântico e dos posicionamentos dos sujeitos implicados no domínio político eleitoral, evidenciando as marcas argumentativas presentes em seus pronunciamentos diacronicamente. Utilizando-se técnicas oriundas do PLN, como a representação estatística de documentos através do cálculo de frequências de palavras; e o uso de SOCs para a identificação de termos-chave no campo do discurso, tal proposta integra as duas dimensões do discurso, quais sejam, uma *linguístico-enunciativa* e outra *discursivo-situacional*.

A dimensão *linguístico-enunciativa* refere-se à organização do discurso em termos de suas propriedades lexicais e de suas relações semânticas, em que se determina a encenação da qual participam os sujeitos, constantemente regulada no próprio processo discursivo, a partir das restrições da dimensão *discursivo-situacional*. Já a dimensão *discursivo-situacional* configura, assim, como a instância em que se produz o acontecimento discursivo, coordenado por uma relação contratual que os sujeitos estabelecem entre

si, como reconhecimento das regras e convenções que organizam o direito às suas enunciações em um dado domínio social, a exemplo do domínio político (Charaudeau, 1996). Assim, a dimensão situacional é vista como base da dimensão linguística e, ao mesmo tempo, constituída por esta.

A análise de um gênero discursivo pressupõe, antes de tudo, a compreensão do conceito do gênero em questão e dos parâmetros que o definem. Emediato (2003, p. 64) compreende o gênero discursivo como um tipo situacional, “*como um produto de uma interação complexa entre as condições de produção dos textos e das variantes textuais e as competências que estariam na base da produção dos discursos sociais*”. Subjaz a esse entendimento o conceito de ancoragem social do discurso, que preconiza que não há produção de texto que não se encontre atrelada às determinações impostas pelo domínio de prática social onde ela se realiza, seja esse religioso, político, jurídico, científico, educativo ou midiático (Charaudeau, 2006; Emediato, 2003). O domínio de prática social surge, então, como “*um determinante maior da natureza comunicacional do discurso*” (Emediato, 2003, p. 65).

Todo ato de linguagem se realiza, dessa forma, em uma situação de comunicação, que determina o espaço da troca linguageira e suas condições de produção, trazendo limitações no seu processo de encenação. O conceito de *situação de comunicação* remete, necessariamente, ao conceito de *contrato de comunicação*. Todo domínio de comunicação propõe a seus parceiros um conjunto de condições nas quais se realiza qualquer ato de linguagem e que normatizam as trocas comunicativas, de modo que os parceiros possam se entender sobre o que constitui a expectativa da troca – uma espécie de contrato de reconhecimento. Tal contrato permite, assim, aos parceiros de uma troca linguageira reconhecerem um ao outro com traços identitários que os definem como sujeitos desse ato – identidade; reconhecerem o objetivo do ato de fala – finalidade; entenderem sobre o que constitui o objeto temático da troca – propósito, e considerarem a relevância das coerções materiais que determinam esse ato – circunstâncias.

No nosso caso, o domínio político, mais especificamente o domínio político eleitoral, determina os rituais linguageiros que podem ser construídos no espaço da política. São exemplos desses rituais a configuração discursiva dos debates instituídos nesse domínio, os papéis ou *scripts* que podem ser desempenhados pelos sujeitos envolvidos de acordo com as expectativas dos seus interlocutores, as apresentações nas reuniões com os comitês de campanha e convenções do partido, a estrutura das declarações na mídia e em comícios, em uma entrevista ou em uma propaganda de campanha e a maneira de abordar temas que constituem as maiores preocupações do eleitorado.

METODOLOGIA

O corpus deste estudo compreende um conjunto de textos proferidos por Dilma Rousseff, ex-presidente da República do Brasil, desde a fase de sua candidatura no primeiro pleito até seu *impeachment*, no segundo mandato. Para efeito de análise, o conjunto de textos abordado foi dividido em séries distintas que possibilitaram a compreensão das diversas apropriações semânticas contextuais nos discursos de Dilma e do estabelecimento de significados contingentes diacronicamente.

A primeira fase da metodologia consistiu na realização de uma análise lexicométrica viabilizada pela identificação e pela extração automática das unidades lexicais (substantivos, adjetivos, verbos, pronomes, sintagmas nominais e nomes próprios) mais frequentes em cada texto analisado, com o

auxílio do *Parser Palavras*¹. A lexicometria é um método estatístico comparativo, que consiste no levantamento da distribuição das unidades lexicais e em comparar ou opor os diversos usos dessas unidades em fragmentos do discurso, com a utilização de ferramentas computacionais. Ela abrange, assim, duas operações principais: (i) a escolha, após a divisão da cadeia textual em unidades que podem ser estudadas; e (ii) a combinação, a comparação ou a oposição das formas textuais identificadas nos discursos que compõem o conjunto da enunciação analisado (Bonnafous; Tournier, 1995; Charaudeau; Maingueneau, 2008).

Na segunda fase foram mapeadas as redes terminológicas – aquelas formadas por termos utilizados pelos sujeitos em processos interacionais em determinado domínio -, que representam a associação entre os termos utilizados por Dilma em sua estratégia argumentativa na dinâmica do debate político. Tais redes são plotadas por meio de um método de análise de coocorrência e avaliadas por meio da contextualização do emprego dos termos e das conformações que elucidam a dinâmica da rede.

Compreendida como um artifício metodológico, a abordagem de redes sociais engloba um conjunto de teorias, conceitos, métodos e modelos das ciências sociais, permeado por diversas perspectivas disciplinares e epistemológicas. A análise de redes sociais é uma abordagem focada na estrutura relacional, que visa a investigar de que forma as relações e os atributos dos atores influenciam o comportamento de cada componente na rede e do todo. Certas unidades básicas compõem as redes. Um ator é uma “unidade discreta individual, corporativa ou social coletiva” (Wasserman; Faust, 1999, p. 17). O *laço relacional* estabelece a ligação entre pares de atores e “são canais de transferência ou fluxo de recursos, sejam eles materiais ou não materiais” (Wasserman; Faust, 1999, p. 4). Um dos objetivos mais importantes da análise de redes consiste na identificação de subgrupos coesos no interior de um conjunto social. Um *subgrupo* coeso é composto por um conjunto de atores que estabelecem relações fortes, intensas, diretas e frequentes (Lazega, 2007). A análise *centro-periferia*, proposta por Borgatti e Everett (2000), auxilia na identificação desses subgrupos. Tal análise pressupõe a existência de dois grandes grupos em uma rede: um grupo de atores fortemente relacionados, denominado de *centro*, e um grupo no qual os atores têm poucos contatos entre si, mas que se ligam ao centro da rede por alguns de seus membros, denominado *periferia*. Nesta região está concentrada a maior parte dos atores da rede. De maneira semelhante, a identificação de *cliques*, subconjunto de membros todos adjacentes uns aos outros que mantêm relações mais próximas entre si, possibilita conhecer quem pertence a um grupo, bem como os tipos e padrões de relacionamentos que os definem e os sustentam como um grupo (Lazega, 2007). Por fim, outra medida importante refere-se à *centralidade de grau* (*degree centrality*) que define a posição de um ator em relação às trocas e às comunicações na rede, considerando-se a quantidade de ligações diretas entre eles (Everett; Borgatti, 2005).

No contexto deste estudo, cada termo representa um ator na rede. Para o mapeamento da rede, foram desenvolvidas algumas rotinas em linguagem de programação *Python*. O algoritmo que possibilita tal mapeamento pode ser descrito pelas seguintes etapas: (i) calcula-se a frequência de cada termo no parágrafo para gerar uma lista de termos; (ii) determina-se a coocorrência entre cada par de termos – entendida como a presença dos dois termos no parágrafo; (iii) atribui-se um peso à coocorrência – esse peso se relaciona inversamente à distância entre os termos no mesmo parágrafo; (iv) sumarizam-se todas as coocorrências em parágrafos distintos com os pesos (atribuindo o peso calculado no passo

¹ Trata-se de um programa, desenvolvido por Eckhard Bick na *Southern University of Denmark*, que incorpora as regras para tratamento morfológico e sintático exigidas pela língua portuguesa, além de realizar a extração dos sintagmas diretamente nos documentos. Uma versão para avaliação está disponível em: <<http://visl.sdu.dk/visl/pt/>>. (Bick, 2000).

anterior a cada ocorrência) e pondera-se pela frequência dos termos no texto; e (v) gera a matriz de coocorrências de um texto a partir dos valores sumarizados.

Os dados gerados pelo algoritmo foram importados para o *software* de análise de redes sociais Ucinet (Borgatti; Everett; Freeman, 2002). Assim, determinaram-se diversas medidas para os termos, como graus de centralidade, coeficientes de agrupamento (cliques) e análise centro-periferia. Esses indicadores forneceram subsídios relevantes para analisar as redes e possibilitaram uma análise comparativa da variação do emprego da terminologia ao longo do tempo e das diferentes conformações assumidas pelas redes.

RESULTADOS

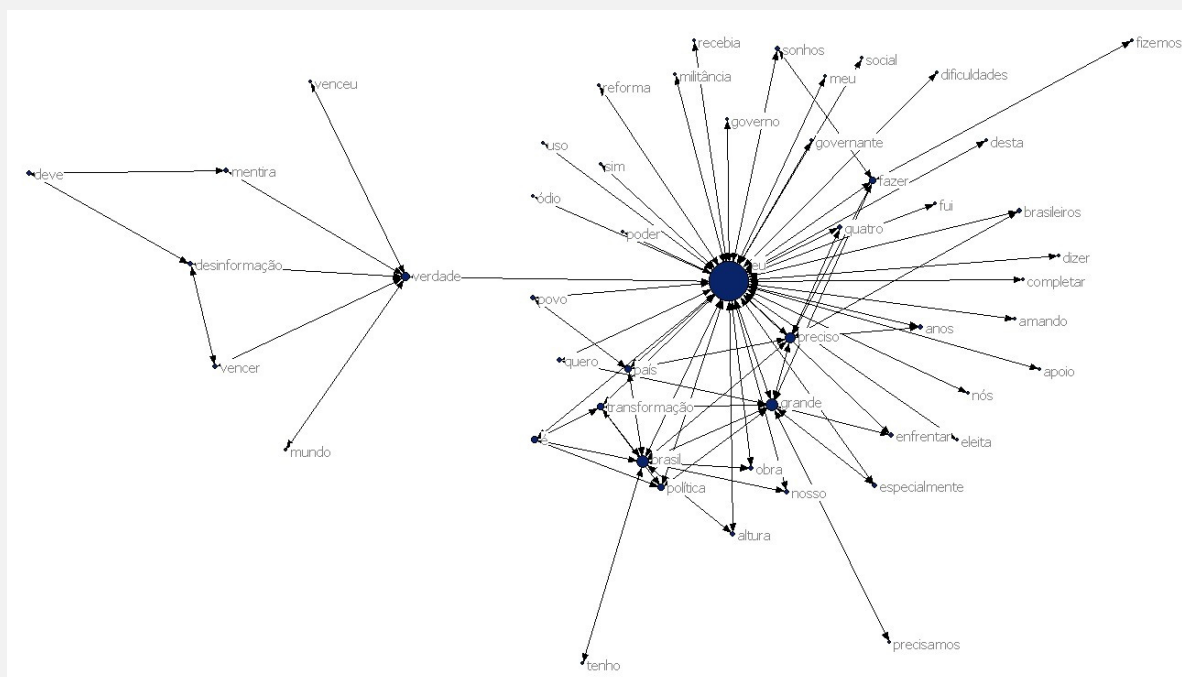
Ao se procurar caracterizar as representações discursivas, logo se percebe que o fazer político é sempre situado e apropriado pragmaticamente em processos interacionais. Desta forma, tentando ajustar o foco sobre a especificidade do domínio em análise, cabe uma descrição da cena enunciativa das eleições presidenciais brasileiras de 2010 e 2014, bem como da candidata Dilma Rousseff.

O cenário das eleições presidenciais de 2010 foi bastante distinto do encontrado nos pleitos anteriores. Pela primeira vez, após a redemocratização do país, Luiz Inácio Lula da Silva não seria lançado como candidato à Presidência da República pelo Partido dos Trabalhadores (PT), pois a Constituição Brasileira não permite o exercício de mais de dois mandatos consecutivos como Chefe do Executivo. Nove candidatos disputaram o cargo presidencial em 2010, dentre eles Dilma Rousseff.

Dilma, representante do governo, é economista e foi Ministra da Casa Civil na gestão do governo do Presidente Lula (2003-2009). Candidata pela coligação “*Para o Brasil Seguir Mudando*”, teve como vice o deputado Michel Temer. Foi candidata pela primeira vez à Presidência da República e indicada por Lula que antecipou o processo de discussão partidária usual nesse estágio. Foi responsável pela criação de programas estratégicos como o *Plano de Aceleração do Crescimento* (PAC) e o programa *Minha Casa, Minha Vida*. Eleita, em 2010, no segundo turno com 46,91% dos votos válidos, alcançou dois fatos inéditos na história política brasileira: ser a primeira mulher presidente do Brasil e exercer o terceiro comando nacional sucessivo de um mesmo partido.

Em 2014, o cenário político-eleitoral do Brasil era preocupante, havia aumentado a apreensão em relação às contas públicas e ao desempenho da indústria no Brasil, e a rejeição popular ao governo de Dilma só crescia. Dilma se reelegeu em 2014, no segundo turno, com 51,6% dos votos válidos e sofreu processo de *impeachment* em 2016, acusada de crimes de responsabilidade fiscal.

Como abordado anteriormente, os textos foram divididos em séries e analisados diacronicamente. O primeiro conjunto de textos refere-se aos discursos de candidatura das eleições de 2010 e 2014. Nessa fase da campanha, Dilma articulou seu discurso em torno das propostas de campanha, pautadas pela agenda dos fatos políticos, em que se busca ampliar o apoio no interior do partido, das alianças estabelecidas com outros partidos, dos movimentos sociais e de outros grupos da sociedade civil em geral. Destaca-se a utilização das formas *Brasil* e suas respectivas derivações, bem como termos correlatos como *país*, *povo*, *população* e *sociedade*, e a recorrência aos termos associados aos projetos e pré-programas de governo, como: *desenvolvimento*, *educação*, *trabalho*, *saúde*, *cultura* e *segurança*.



Obs.: Rede com 50 termos apresentados segundo medida de centralidade de grau

FIGURA 2: Rede terminológica – Discurso de candidatura 2º pleito

Fonte: Dados do estudo gerados com auxílio da ferramenta Ucinet

Em relação à segunda série, que engloba os discursos de posse, há uma remissão ao governo anterior, principalmente porque Dilma representava o terceiro mandato de Lula, em 2010, uma fala permeada por agradecimentos a aliados e consolidação de alianças futuras, e uma projeção de um país desejado que seria viabilizado pelo novo mandato. Os discursos desta série se assemelham ao da série anterior, uma vez que Dilma fez do discurso de posse uma lista de obras e projetos em andamento e que precisam melhorar ou continuar – estratégia pouco comum em discursos dessa natureza, dado o contrato de comunicação político eleitoral. Assim, destacam-se como termos centrais nas redes: *investimento, desenvolvimento, compromisso, crescimento, mudança*, formando subgrupos com os termos *economia, política, reforma e educação*, além dos verbos que projetam ações no futuro, tais como: *ampliar, avançar, continuar, fazer e vamos* (ver Figuras 3 e 4). Destaca-se ainda o agrupamento formado pelos termos *Brasil, brasileiros, povo, governo, país e vamos*, evidenciado pela análise centro-periferia, e que remetem a uma reafirmação do acordo estabelecido entre a candidata e seu eleitorado.

CONCLUSÕES

A metodologia proposta propicia a construção de uma rede destinada a expressar as relações entre os termos, considerando-se o contexto particular onde aparecem, e possibilita o estabelecimento de relações inusitadas entre eles. Evidencia-se que o aprofundamento de estudos de redes terminológicas possibilita descrever e explicar melhor as dinâmicas e multifacetadas relações sociais estabelecidas entre diferentes sujeitos, para o que abordagens, como a empregada nesta comunicação, que combinam técnicas oriundas do processamento (automático) da linguagem natural (PLN), como a representação estatística de documentos através do cálculo de frequências de palavras; e o uso de sistemas de organização do conhecimento (SOC/KOS) para a identificação de termos-chave no campo do discurso, são potencialmente benéficas.

Por fim, a formalização explicativa, empreendida pela abordagem aqui proposta, através da análise de alguns processos discursivo-argumentativos que são constitutivos da maneira como os sujeitos efetivam os discursos sociais, permite melhor compreender as configurações e remanejamentos das trocas discursivas em domínios dinâmicos, tal como o político, e, conseqüentemente, reduzir o gap entre a formulação conceitual e sua representação em diferentes dispositivos no âmbito da organização do conhecimento.

REFERÊNCIAS BIBLIOGRÁFICAS

- Bick, E. (2000). *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press.
- Bonnafeous, S., & Tournier, M. (1995). Analyse du discours, lexicométrie, communication et politique. *Langages*, 29(117), 67-81.
- Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social networks*, 21(4), 375-395.
- Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.
- Charaudeau, P. (1996). Para uma nova análise do discurso. Carneiro, A. D. *O discurso da mídia*. Rio de Janeiro: Oficina do Autor, 5-43.
- Charaudeau, P. (2006). *Discurso político*. Contexto.
- Charaudeau, P., & Maingueneau, D. (2008). *Dicionário de análise do discurso*. Contexto.
- Emediato, W. (2003). Os gêneros discursivos como tipos situacionais. Mari, H et al. *Análise do discurso em perspectivas*. Belo Horizonte: Nad/Fale/UFMG, 63-72.
- Everett, M. G., & Borgatti, S. P. (2005). Extending centrality. Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2005). *Models and methods in social network analysis*. Cambridge university press, 57-76.
- Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Digital Library Federation, Council on Library and Information Resources.

Lazega, E. (2007). *Réseaux sociaux et structures relationnelles: «Que sais-je?»*. Presses universitaires de France.

Marciniak, M., & Mykowiecka, A. (2009). Aspects of Natural Language Processing. *Lecture Notes in Computer Science*, 5070.

Souza, R. R., Tudhope, D., & Almeida, M. B. (2012). Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. *Knowledge organization*, 39(3), 179-192.

Venâncio, L. S. (2014). *Encenações linguageiras, jogos argumentativos e redes terminológicas nas eleições presidenciais brasileiras de 2010: a representação da informação em domínios dinâmicos*. Tese de doutoramento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Minas Gerais, Brasil.

Wasserman, S., & Faust, K. (1999). *Social network analysis: methods and applications*. Cambridge: University Press.