
Variação em línguas românicas: ações do projeto VariaR como contributos de ciência aberta e cidadã

Variation in Romance Languages: Actions of the VariaR Project as Contributions to Open and Citizen Science

Variation en langues romanes : les actions du projet VariaR comme contributions à la science ouverte et citoyenne

Marcia dos Santos MACHADO VIEIRA Vanessa MEIRELES

 <http://interfas.univ-tlse2.fr/reflexos/1325>

Marcia dos Santos MACHADO VIEIRA Vanessa MEIRELES, « Variação em línguas românicas: ações do projeto VariaR como contributos de ciência aberta e cidadã », *Reflexos* [], 6 | 2023, 19 avril 2023, 19 avril 2023. URL : <http://interfas.univ-tlse2.fr/reflexos/1325>

CC BY

Variação em línguas românicas: ações do projeto VariaR como contributos de ciência aberta e cidadã

Variation in Romance Languages: Actions of the VariaR Project as Contributions to Open and Citizen Science

Variation en langues romanes : les actions du projet VariaR comme contributions à la science ouverte et citoyenne

Marcia dos Santos MACHADO VIEIRA Vanessa MEIRELES

Introdução

Variação em línguas românicas: *inCorpora*

Portal digital de estados de coisas em português e línguas românicas a variar e ensinar

Introdução

Language norms are dynamic conventions that change over time. As Amato et al. (<https://www.mdpi.com/2226-471X/7/1/66#B1-languages-07-00066>) (2018, p. 8260 (<https://www.mdpi.com/2226-471X/7/1/66#B1-languages-07-00066>)) point out,

Broadly speaking, a convention is a pattern of behavior shared throughout a community and can be defined as the outcome that everyone expects in interactions that allow two or more equivalent actions (e.g., shaking hands or bowing to greet someone).

Norms change because of language use and also as a result of various external factors. The driving forces of norm change can be either formal institutions, such as language academies, or informal institutions, such as mass media, but change can also be unregulated. (CUENCA, 2022, p. 1)

A ‘diasystem’ can be constructed by the linguistic analyst out of any two systems which have partial similarities (it is these similarities which make it something different from the mere sum of two systems). But this does not mean that it is always a scientist’s construc-

tion only: a ‘diasystem’ is experienced in a very real way by bilingual (including ‘bidialectal’) speakers [...]. (WEINREICH, 1954, p. 390)

- 1 Apresentamos brevemente movimentos, trilhas e horizontes relativos ao objetivo do projeto franco-brasileiro VariaR – Variação em línguas Românicas/UPVM-UFRJ¹ – de incorporar, numa rede de ações e interações, sujeitos (pesquisadores, professores, estudantes e outros interessados em dados (socio)linguísticos), *corpora* representativos de usos dessas línguas e de suas variedades e variações, bem como informações sobre convergências e divergências entre as gramáticas delas, sobre os textos em que essas gramáticas emergem e sobre as comunidades que delas se servem para se expressarem diversamente.
- 2 A meta com essa rede é, em linhas gerais, viabilizar a comparação das línguas românicas em uso no mundo, o amplo conhecimento de seus diassistemas e o trabalho com eles em espaços de ciência e educação respaldado por dados da realidade heterogênea e pluricultural e por textos que espelham essa realidade. Tratamos, particularmente, de ações planejadas para a construção do portal digital *inCorpora*, por meio do qual o acesso aberto às coleções de (meta)dados e às descrições e informações conexas pretendidas se dará e mediante o qual a equipe do projeto VariaR pretende realizar e fomentar pesquisas de estabilização, variação e mudança linguísticas.
- 3 Atemo-nos, particularmente, a um dos empreendimentos no movimento de configurar o sistema de informações e de coleções desse portal: o *Portal digital de estados de coisas em português e línguas românicas a variar e ensinar*, um subprojeto que conta com apoio financeiro brasileiro (da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro/Cientista do Nosso Estado 32/2021 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico/MCTI/FNDCT² 18/2021).
- 4 Esse empreendimento integra o plano de interesses científicos e ações em linguística teórica, descritivo-comparativa e aplicada do projeto VariaR. O portal *inCorpora* busca reunir e partilhar, em acesso aberto e com respeito às devidas licenças/autorizações, coleções de (meta)dados do Português e de outras línguas românicas em uso em práticas sociais escritas, pelo menos a princípio. Possivelmente, um dia poderá vir a abrigar também coleções oriundas de práticas sociais

orais ou faladas. Planejamos, mais precisamente, a construção e curadoria de um portal digital que contenha, no início, *corpora* escritos, bem como usos e descrições de predicação verbal e temas correlacionados. A diretiva é resolver o problema de acesso a dados e descrições qualificadas de expressão de estados de coisas³ via predicações verbais em línguas românicas, especialmente em espaços de trabalho com português língua (não-)materna. Consideramos que há usos assim como convergências e divergências a detectar com base na comparação entre línguas românicas e variedades destas que podem alimentar bases de trabalho nacionais e internacionais relativas a práticas comunicativas e interacionais diversas. Falta um repositório temático ou um grande banco de coleções de textos e de dados que viabilize isso.

- 5 Prevedemos que um portal dessa ordem poderá vir a colaborar para pesquisas de diversos fenômenos linguísticos (além dos associados à predicação e a predadores), para o ensino de uma língua românica atento à realidade de seus usos e às suas descrições empíricas, para a conscientização das comunidades sobre o que é estável e o que varia a depender de condições funcionais e socioculturais, para a percepção da configuração identitária de comunidades e indivíduos que se exprimem via linguagem, para armazenamento e salvaguarda de estágios de manifestações das línguas e de aspectos socioculturais associados, bem como, por tabela, para educação patrimonial imaterial. Tem, portanto, contributos a oferecer à ciência e educação abertas e cidadãs.

Variação em línguas românicas: *inCorpora*

A gestão de dados para pesquisa é uma atividade estratégica que favorece a integridade, a qualidade, a reprodutibilidade da pesquisa, a memória institucional e o reuso dos dados pelo próprio grupo de pesquisa e/ou por outros grupos. (FUNDAÇÃO OSWALDO CRUZ, 2020, p.8)

- 6 Norteado por uma política de tecer uma rede de investigadores, pesquisas, *corpora* e descrições das línguas românicas em uso no mundo, a coordenação do projeto VariaR vem, desde 2019, promo-

vendo interações e ações para a consecução desse propósito e, mais recentemente, vem trabalhando no sentido de tornar coleções de dados (socio)linguísticos (coletados ou produzidos via atividade científica) e coleções de textos objetos de: exposição e valorização da heterogeneidade linguístico-cultural; preservação de patrimônio imaterial; acesso aberto, por qualquer interessado em qualquer canto do mundo; (re)uso em educação aberta e cidadã e em outras esferas de atuação e impacto na sociedade.

- 7 Nesse sentido, o projeto VariaR procura dar atenção a estudos de contatos multilíngues/multidialectais e a investigações de fenômenos linguísticos (fonético-fonológicos, morfológicos, sintáticos, lexicais e textuais-discursivos) em variação, estabilização ou mudança. Perspectiva desenvolver-se à luz da Sociolinguística e de suas interfaces (entre as quais, a interface socioconstrucionista, conforme caracterizada por MACHADO VIEIRA e WIEDEMER, 2019 e 2020) ou à base do diálogo com outras abordagens ou mesmo noutra abordagem de estabilidade e variação. Programa a análise descritivo-comparativa dos graus de diassistematicidade (HÖDER, 2012) detectáveis, na primeira fase, em espanhol, francês, italiano, português e romeno e entre as variedades dessas línguas, especialmente em relação à temática de predicações verbais e a temas gramaticais conexos. Prevê, no seu desenrolar, a possibilidade de expansão para as demais línguas românicas e outros fenômenos linguísticos.

Diasystematic links and dia-elements constitute a network through which two language systems used within a multilingual speaker group are interconnected. The degree to which two varieties in contact participate in the common diasystem depends, of course, on their typological similarity: closely related and typologically similar languages can more easily develop a high degree of diasystematicity – i.e. the common intersection of their systems is larger – than more distant languages, which retain a larger proportion of idiosyncrasies in their systems [...]. (HÖDER, 2012, p. 10)

- 8 Nossa hipótese principal é a de haver uma configuração gramatical diassistemática baseada em diaconstruções que permitem usos linguísticos interligados nessas línguas e, assim, constituem padrões construcionais, com atributos formais (lexicais e morfossintáticos) e funcionais (cognitivos, semânticos, discursivos, pragmáticos e so-

ciais), compartilhados. A ideia é, então, mapear tais padrões construcionais e também as idioconstruções que são particulares à gramática de cada língua. Entendemos que a hipótese de existência de diaconstruções e idioconstruções tem lugar, mesmo quando se comparam variedades do Português. E isso é válido até mesmo quando a comparação envolve variedades nacionais, em que costuma sobressair a perspectiva do contraste/da diferença.

De fato, vários campos da pesquisa linguística já reconhecem que não há como considerar o português de Portugal e do Brasil a mesma língua. É verdade que o que define as relações linguísticas internas e externas não é somente o nome, mas também passa por ele. É por isso que, com base na reflexão dos pesquisadores citados, sigo-os ao defender a expressão *português brasileiro*.

Chamar apenas “português”, na perspectiva histórica da língua, é um anacronismo, ou seja, não corresponde aos fatos históricos, visto que desconsidera as noções de desenvolvimento independente e contínuo que a língua portuguesa teve no Brasil, especialmente após o século XIX. Ao mesmo tempo, chamar apenas de “brasileiro” também evidencia um anacronismo, pois desconsidera toda a herança linguística que foi trazida para o Brasil, que é inegavelmente portuguesa. O nome “português brasileiro” representa, assim, ruptura e continuidade. É um nome apropriado não apenas nos contextos de pesquisa acadêmica, mas uma importante ferramenta para que os falantes se reconheçam diante de sua língua. Falamos português, mas o português do Brasil. (OLIVEIRA JR., 2022, <https://www.roseta.org.br/2022/05/11/portugues-ou-brasileiro-qual-e-o-nome-da-nossa-lingua/>)

- 9 Em termos gerais, o projeto VariaR (cf. MACHADO VIEIRA e MEIRELES, 2022b, p. 211) tem como horizonte assentar-se em corpora de natureza diversa que possibilitem a investigação em diferentes linhas de trabalho para ir ao encontro das necessidades e potencialidades das ciências humanas (digitais ou não) e de demandas da sociedade em geral: cartografia e documentação, tradução, ensino, interpretação ou dublagem, comunicação ou interação/rotina (socio)discursiva em diferentes áreas do conhecimento, softwares e algoritmos que promovam a gamificação do conhecimento sobre línguas românicas, entre outros. Este projeto desenvolve-se à luz das

teorias que permitam lidar com variação, estabilização, mudança linguística e contacto multilingue e multidialetal.

- 10 Outro movimento da coordenação do projeto VariaR ocorre no sentido de criar condições para que, via modos de trabalho colaborativo, seja possível disponibilizar dados e textos dessas línguas a toda a equipe e, tomados os devidos cuidados legais e éticos, dar acesso aberto para que se comparem as descrições de suas gramáticas (encaradas como diassistemáticas, conforme HÖDER, 2012, 2014) e os textos/as textualidades em que estas se apresentam. Dessa maneira, esperamos contribuir para o conhecimento dessas línguas e a gestão desse saber em materiais e espaços de ensino delas como línguas maternas ou não maternas assim como de desenvolvimento de infraestrutura tecnológica. Cuidamos, além disso, de fomentar: o direito da sociedade ((não-)científica, escolar, em geral) à informação qualificada sobre a realidade multilinguística e multidialetal; a sensibilização no olhar (que, esperamos, passe a ser justo, não estereotipado, respeitoso e sem preconceito) desta e de seus usuários; a economia sustentável de coleções de (meta)dados linguísticos, para que coleções de dados/textos cujo destino geralmente tem sido o de reserva pessoal ganhem visibilidade, otimização e (re)uso em novas pesquisas e/ou em consultas; sua potencial repercussão quanto à construção, governança e atualização de sistemas de informações, atributos semânticos e ontologias na área de linguagens, à administração de negociações entre pessoas físicas e/ou jurídicas que se dão via interações linguísticas, a marketing, a turismo, à inteligência artificial, à tradução automatizada, a interações interculturais em ambiente real, virtual e híbrido/virtual imersivo (como previsto no âmbito do que atualmente é chamado de metaverso); a inovação tecnológica na área de linguagem, mediante a captação de esforços de sujeitos de áreas afins e de ferramentas/algoritmos para a consecução de planos, soluções e espaços (físicos, virtuais ou digitais) que levem à visibilidade dessas coleções temáticas.

Ciência Aberta é um termo guarda-chuva que abrange as práticas de abertura de dados seguindo os princípios da transparência e colaboração na Ciência. Atualmente, o tema é estratégico para diversos países, que já a adotam como política pública para que organizações de pesquisa passem a incorporá-la gradativamente como princípio orientador de suas atividades.

A abertura de dados passou a ser vista como estratégica por fomentar o desenvolvimento de infraestruturas e tecnologias que estimulam sua disponibilização e por facilitar o intercâmbio e a interoperabilidade desses dados entre diferentes sistemas.

(...)

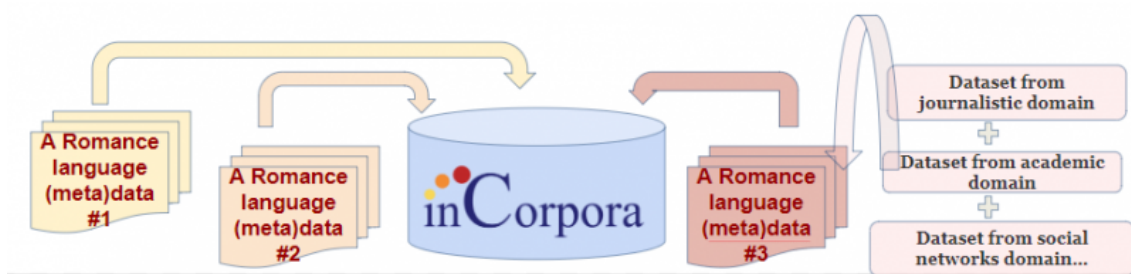
Além de trazer diversos benefícios ao processo de comunicação científica, como maior celeridade, confiabilidade e redução de custos, os dados depositados em repositórios são citáveis, configurando-se então como produções científicas legítimas e reconhecidas. Desta forma, eles contribuem para aumento da visibilidade da produção científica nacional. (<https://www.rnp.br/noticias/o-que-e-ciencia-aberta-e-como-ela-pode-facilitar-vida-de-cientistas>)

- 11 No que diz respeito ao trabalho colaborativo, temos feito movimentos – especialmente por meio de reuniões, colóquios, seminários e ateliês – para envolver pesquisadores de diferentes línguas românicas e para integrar ações relativas a diferentes frentes desse trabalho. Já contamos com pesquisadores da França, Brasil, Itália, Portugal, Romênia, Alemanha, Guiana, Coreia do Sul atuantes na equipe do projeto VariaR⁴. Em nosso horizonte, estão previstas estas frentes de trabalho a congregá-los: definição das fontes de pesquisa, dos tipos de coleções de documentos das línguas românicas a reunir ou levantar e dos modos de o fazer atentos aos princípios FAIR⁵ (*Findable, Accessible, Interoperable, Reusable*), a um programa de standardização de (meta)dados que promova interoperabilidade, às licenças de compartilhamento e aos instrumentos de direito de autoria de conteúdos e de direitos conexos (cf. ROBIN, 2022); reunião, criação e/ou organização de coleções de objetos/documentos para consulta ampla (dados linguísticos, textos, descrições, por exemplo); construção de um repositório/portal digital para acolher tais objetos – intitulado por nós, até o momento, de *inCorpora*⁶ – e para permitir, em alguma medida, interoperabilidade multiusuário e multissistêmica e leitura por humanos e máquinas; planejamento de um sistema de descritores dos metadados desses objetos/documentos linguísticos; tratamento, salvaguarda e exploração de dados linguísticos, textos e descrições no portal e mesmo em outros ambientes (museológicos, pedagógicos, virtuais, bibliográficos, entre outros) e por usuários diversos (outros cientistas da área de Letras e Linguística, profissionais de outras áreas do saber, professores, estudantes, graduandos em formação em

cursos de Bacharelado ou de Licenciatura e, enfim, qualquer interessado em diversidade (socio)linguística).

- 12 O **inCorpora** é um espaço virtual ligado ao projeto VariaR que visa a reunir ou a acolher coleções de (meta)dados/textos oriundos de diferentes fontes e domínios discursivos (jornalístico, acadêmico, redes sociais, por exemplo), estruturadas e parametrizadas para permitirem o mapeamento das variações, mudanças e estabilizações em línguas românicas, bem como do que há de interlíngua.

Figura 1: O portal digital **inCorpora**

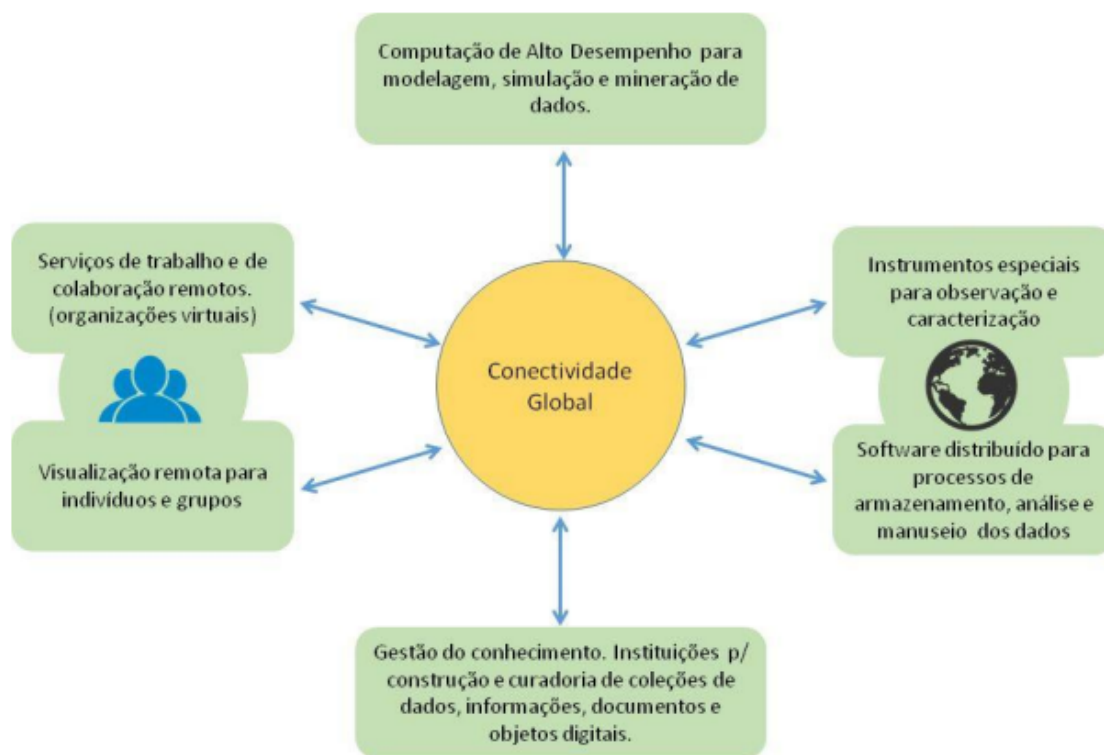


Fonte: MACHADO VIEIRA e MEIRELES, 2022 – The digital portal - <https://twitter.com/marciamv2/status/1599852404802957312> Acesso em: 03/01/2023.

- 13 Entre os empreendimentos para povoar o portal **inCorpora** e dar visibilidade a coleções de dados de línguas românicas, estamos desenvolvendo o projeto **Portal digital de estados de coisas em português e em línguas românicas a variar e ensinar**, desde o início de 2022, com foco no português perspectivado em comparação com outras línguas românicas (especialmente, o francês e o espanhol, mas também italiano e romeno, futuramente) e, para tanto, contando com apoio financeiro de duas agências de fomento brasileiras (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro/FAPERJ e Conselho Nacional de Desenvolvimento Científico e Tecnológico/CNPq), bem como contando com uma equipe de pesquisadores (docentes e discentes) da área de Ciência da Computação da UFRJ (Universidade Federal do Rio de Janeiro⁷) e do CEFET/RJ (Centro Federal de Educação Tecnológica Celso Suckow da Fonseca⁸). A ideia é a de que, com o tempo, também se somem outros projetos com foco em cada uma das demais línguas e sempre perspectivando a comparação de línguas românicas.

- 14 Com isso, esperamos consolidar e adensar uma rede de expertises e ações interinstitucionais e estratégicas em torno das línguas sob nossa atenção, em termos de diretrizes teóricas, ciberinfraestruturais, metodológicas e descritivas (minimamente) comuns e, por conseguinte, capazes de permitir comparação e mensuração do grau de diassistematicidade das línguas românicas. E esse horizonte demanda expertise em tecnologias de informação e comunicação (TIC), em ciberinfraestrutura (CI).

Figura 2: Componentes de ciberinfraestrutura



Fonte: SIMÕES; STANTON; RIBEIRO; GRIZENDI; MACHADO; CIUFFO - Figura 1. Componentes de ciberinfraestrutura ⁹, p. 2, https://www.rnp.br/arquivos/ciberinfraestrutura_rnp.pdf : “resume como esses componentes da CI se relacionam para habilitar um ambiente controlado, seguro, abrangente, de acesso simplificado e econômico, de recursos compartilhados, para o desenvolvimento de pesquisa e educação.” – Acesso em: 03/01/2023.

- 15 Em nosso horizonte está a possibilidade de contribuir, por meio da configuração de um portal digital, para um melhor conhecimento das variedades do português nos espaços em que é língua materna e além destes e, por extensão, da área linguística lusófona de forma mais geral em espaços das línguas românicas, bem como para um melhor

conhecimento de variedades das línguas românicas (cf. MACHADO VIEIRA e MEIRELES, 2022a e 2022b). Queremos fazê-lo detectando e descrevendo um terreno comum sobre as variações dialectais contemporâneas dessas línguas e proporcionando um ambiente virtual para reunir informações sobre as generalizações alcançadas nesse terreno, bem como as especificidades dessas línguas e das comunidades que as usam.

Portal digital de estados de coisas em português e línguas românicas a variar e ensinar

- 16 Nesse projeto em particular, estão em foco tipos de construções de predicação verbal e predicadores a serem observados em *corpora* escritos de variedades brasileiras do português, bem como deles em comparação com dados escritos de variedades das línguas românicas (a começar pelo francês e pelo espanhol e futuramente também a envolver italiano e romeno), perspectivando ciência da estabilização, variação, mudança assim como ensino de português em espaços de língua materna e não materna. Como ainda está no primeiro ano de desenvolvimento, a fase do ciclo de vida de (meta)dados relativa a conceptualização do perfil das coleções, planejamento ou testagem de rumos científicos, ciberinfraestruturais e tecnológicos é a que mais tem tomado nossa atenção.
- 17 São estas as primeiras frentes de análise de predicação verbal planejadas para povoar o portal *inCorpora* com resultados de descrições empíricas: a) construções de perspectivação passiva de um estado de coisas (ter/sofrer menção – *tener/sufrir mencion*, *avoir mention* – ser mencionado/*être mentionné* – mencionar-se/se *mentionner/mencionarse*); b) construções de perspectivação perceptual/modal de um estado de coisas (dar uma olhada – dar uma vista dólhos – *lancer un regard/echar un vistazo*; ficar roxo de irritação/*être très en colère/se volver morado de irritación* – irritar-se muito); c) construções de representação de um estado de coisas (fazer-se/tirar de tonto/*hacerse el tonto* ou fazer-se de vítima/*jouer la victime*; trazer sugestão/sugerir); d) construções de prospecção/futuridade, ou seja, de perspectivação de estados de coisas com projeção futura (vai fazer,

va faire; quer fazer, *veux faire*; estar a ponto de fazer, *être sur le point de faire*). Queremos investigar como são conceptualizados estados de coisas segundo diferentes comunidades e contextualidades (domínios e gêneros discursivos; espaços de língua materna ou não; atividades de ensino, tradução, comunicação).

- 18 Prevemos enfrentar, sob análises multivariadas e, possivelmente, colostrucionais/colocacionais, o problema da diassistematicidade com base na detecção de padrões colocacionais e construcionais de predicação estáveis e comuns (diaconstruções) às línguas românicas consideradas e suas variedades e na detecção de seus padrões particulares (idioconstruções), bem como de padrões construcionais internos às variedades que se alinhem por similaridade (aloconstruções/variantes construcionais) numa área na rede construcional (metaconstrução) resultante do mecanismo cognitivo de analogia e, então, do processo de neutralização dos atributos que, noutros contextos, servem para os diferenciar. Tencionamos lidar com a relação entre essas generalizações teórico-descritivas sobre predicações de perspectivação e representação em espaços de trabalho com Português como língua (não-)materna.
- 19 Projetamos reunir textos das línguas românicas de fontes de diferentes países ou regiões/localidades de um país. Queremos contar, sempre que possível, com textos de mais de um gênero textual e mais de um domínio discursivo. A depender da configuração das coleções de textos e de seus metadados, estimamos, no caso de investigações de predicações verbais, explorar variáveis ligadas ao parâmetro de contextualidade (condições discursivas, cognitivo-pragmáticas e sociais; atos de fala; movimentos retóricos num texto; temáticas; gêneros e domínios discursivos) e o complexo de combinações de construções de configuração formal lexical ou morfossintática em construções de configuração formal textual, com base na postulação de que construções lexicais e morfossintáticas se combinam em construções textuais-discursivas (de configuração formal suprassentencial). Consideramos construção textual-discursiva como construto teórico de representação da rede de construções que é referenciada como Gramática de Construções (cf. WIEDEMER e MACHADO VIEIRA, 2022).

- 20 Esses rumos impactam em decisões relativas a conceptualização e planeamento das coleções de (meta)dados que a equipe do projeto VariaR pretende produzir para o portal. Demandam, por exemplo, decisões coletivas da equipe no sentido de um conjunto de entidades, atributos e categorias para estruturar ontologicamente o sistema de informações do portal.
- 21 Em termos de gestão de *corpora* e (meta)dados destes, prevemos que, além da equipe do projeto VariaR, usuários verificados possam vir a fazer *upload* de arquivos com coleções de dados, informar sobre os metadados desses dados a partir de parâmetros e descritores configurados pela equipe da área de tecnologia de informação e computação e, assim, postar/depositar *corpora* e colaborar para que outros os reutilizem ou simplesmente os acessem (haja vista demandas atuais de acessibilidade aos dados, como o alerta feito pela revista Nature em 2016 <https://www.nature.com/articles/nmeth.4026> sobre a política de informar se e como os leitores podem acessar os dados subjacentes à pesquisa reportada). Também consideramos que *corpora* compartilhados pela equipe do Projeto VariaR podem ser usados por outros usuários, externos à equipe, para realizar novas análises, metanálises ou investigação de outra ordem. Podem, ainda, servir para não especialistas em Linguística simplesmente fazerem consultas ou sanarem curiosidades sobre as línguas românicas. Assim sendo, é preciso cuidar da gestão de conteúdos, considerando, entre outros aspectos, proveniência dos dados, licenças de uso/acesso, formatos e padrões de ontologia convencionalizados na área, infraestrutura de preservação digital a longo prazo, infraestrutura e protocolos de curadoria de conteúdos.
- 22 Afinal, sabemos que metadados (estratégicos, técnicos ou operacionais) bem descritos e gerenciados promovem a confiança nos dados/conteúdos assim como mapeamento e mensuração precisos e qualificados na comparação das línguas, colaboram para otimizar o tempo de implementação técnica no sistema de informações do portal e de consulta ao portal, contribuem para o acesso devido à linhagem dos dados e favorecem o acesso aberto e a recuperação de informações mesmo a longo prazo e mesmo entre os idealizadores do portal. Os metadados estratégicos dizem respeito aos conteúdos e áreas temáticas de interesse, aos sujeitos/papéis a operarem no portal, às condições (padrões ou restrições) de atuação destes e dos

dados/textos, as diretrizes de citação/acreditação; as potenciais repercussões dos conteúdos, à concepção e governança das coleções e de suas conexões. Os metadados técnicos relacionam-se, por exemplo, a detalhes de descrição das coleções dos dados e dos sistemas que as armazenam (data de criação, nomes e propriedades de identificação e localização precisas), à configuração de nível e via de segurança (usualmente mediante replicação dos (meta)dados em locais distintos), a ferramentas/recursos, protocolos e processos que interrelacionam coleções, dados e entidades dentro e entre sistemas. Já os metadados operacionais descrevem detalhes do trabalho tecnológico, do processamento e acesso aos dados, bem como medidas de controle e registros de erros, relatórios de acesso à consulta, frequência e tempo de execução, provisões de *backup* e recuperação de desastres, definição de agendas e ciclos de vida, atualização, *backup* e curadoria de conteúdos.

- 23 Quanto à área de configuração do portal em termos de inteligência artificial, objetivamos uma estruturação que dê margem para funcionalidades tais como: marcação de parte(s) do discurso; extração de palavras-chave; listagem de combinações típicas, sinônimos, antônimos e traduções; garimpagem e localização de exemplos de uso.
- 24 Para o exame de dados linguísticos, prevemos estas funcionalidades: identificação de relações gramaticais entre palavras; pesquisa de colocações e combinações de palavras (em construções gramaticais e construções textuais-discursivas); análise da tendência das expressões ao longo do tempo; análise sobre subcorpora.
- 25 O portal vem sendo concebido segundo possibilidades das novas tecnologias e de estratégias colaborativas para reunir amostragem representativa de nosso repertório patrimonial, arquivá-lo com segurança e com base em identificação persistente, interrelacioná-lo a outros bancos de dados que primam por interoperabilidade e preservá-lo para outras e diversas consultas por longo tempo, na linha do programa de standardização FAIR. Iniciamos, em 2023, o processo de testagem das tecnologias mais compatíveis aos intentos de gerenciamento das coleções de dados, metadados, descrições e conteúdos do portal, tendo em vista estes rumos: sistema de banco de dados personalizado e/ou sistema de repositório, talvez via

- DSpace (<https://dspace.lyrasis.org/>), de olho, por exemplo, no princípio de interoperabilidade por máquinas e por humanos.
- 26 A essência do portal *inCorpora* é, em de modo geral, a de que venha a ser uma ferramenta digital para a promoção de informações qualificadas sobre as línguas românicas no mundo (espanhol, francês, português estão entre as mais usadas segundo Ethnologue - Languages of the World - <https://www.ethnologue.com/guides/ethnologue200>), a salvaguarda de registros e amostras de dados de uso autêntico delas em diversas contextualidades (já que as línguas e as sociedades são dinâmicas), a partilha solidária e o reuso sustentável e otimizado de coleções de dados/textos, além dos usos nas investigações originais que motivaram suas constituições, a visualização de como as línguas românicas e suas variedades efetivamente funcionam e que interrelações mantêm em decorrência de seu parentesco genealógico, da história de seus contatos e de uma atualidade em que o mundo está hiperconectado física e virtualmente.
- 27 Também prevemos cooperar com outros linguistas e com outros profissionais, como os da área de Ciência da Informação e Tecnologia que se interessam por lidar com desafios de inovação e solução tecnológica relativos às redes temáticas de implementação das chamadas Humanidades Digitais, entre os quais os trazidos pela área de Letras e Linguística (uma das áreas, segundo o manifesto das Humanidades Digitais <https://humanidadesdigitais.org/manifesto-das-humanidades-digitais/> e o movimento GO FAIR Brasil <https://www.go-fair-brasil.org/>). De acordo com o manifesto das humanidades digitais, “A opção da sociedade pelo digital altera e questiona as condições de produção e divulgação dos conhecimentos.” E isso também repercute em interações e ações transdisciplinares para prover experimentações, dispositivos, métodos e perspectivas inovadores e condizentes às novas demandas e esferas de atuação na sociedade.
- 28 Vemos, ainda, o potencial de colaborar, em alguma medida, com o projeto Plataforma da Diversidade Linguística Brasileira (<https://www.museudalinguaportuguesa.org.br/projeto-plataforma-diversidade-linguistica-brasileira/>), possivelmente via terreno da interoperabilidade tecnológica e semântica prevista em ambos os projetos (conforme descrito em MACHADO VIEIRA e BARBOSA, 2022),

respeitando-se, naturalmente, as especificidades e contribuições à sociedade particulares deles.

- 29 Como estimado em MACHADO VIEIRA e BARBOSA (2022) e em MACHADO VIEIRA; BARBOSA; FREITAG; BORGES; MEDEIROS (2022), também miramos muitos contributos do projeto (socio)linguístico franco-brasileiro VariaR para ciência e educação abertas e cidadãs, para a ciência da informação e tecnologia, para a ciência de dados, para a ciência do direito e da ética (por conta dos aspectos legais que emergem em novas demandas de acesso aberto), para as ciências da sociedade, da cultura, da comunicação, das humanidades digitais, entre outras. Nosso prognóstico é o de que tais contributos extrapolam o espaço franco-brasileiro em que foi iniciado o projeto VariaR, como já ganha feição a partir da adesão de pesquisadores atuantes noutros países.
- 30 Enfim, estão no horizonte diversas trilhas para maximização de potenciais de esferas de ciência e educação: pôr em prática ou promover educação patrimonial; construir parcerias para isso; auxiliar na formação ou conscientização de pessoas, professores, profissionais; prover subsídios para as redes de implementação temáticas das Humanidades Digitais; contribuir para a confecção de materiais didáticos atentos à realidade dos usos das línguas românicas no mundo, ao respeito às diversas comunidades que usam um idioma em diferentes contextualidades e práticas interacionais, aos direitos dessas comunidades de ciência de usos além dos convencionalizados ou estandardizados porque tidos como exemplares a serem seguidos; dar visibilidade à área de Letras e Linguística; promover conexões com outras áreas do saber; evidenciar que nossa expressão ou construção de identidade em qualquer área do conhecimento se vale de língua, de linguagem verbal e não-verbal; desfazer crenças equivocadas, visões estereotipadas de sujeitos, comunidades e variedades/variações de línguas; convocar mais pessoas da comunidade científica ou não-científica a se engajarem no programa de integração de aspectos da cultura geral à cultura digital de previsibilidade de acesso a coleções de dados autênticos, de partilha solidária e respeitosa destes e de informações sobre estes, de coaprendizagem e inovação com impacto social regulado por demandas sociais, boas práticas metodológicas, expertises, parcerias solidárias e desfechos positivos.

CUENCA, Maria-Josep. “Language Norm and Usage Change in Catalan Discourse Markers: The Case of Contrastive Connectives”. *Languages* 7: 66. <https://doi.org/10.3390/languages7010066>.

FUNDAÇÃO OSWALDO CRUZ. Presidência. Vice-presidência de Ensino, Informação e Comunicação. Política de gestão, compartilhamento e abertura de dados para pesquisa: princípios e diretrizes. Rio de Janeiro, 2020. 19 p.

HÖDER, Steffen. “Multilingual constructions: a diasystematic approach to common structures”. In: BRAUNMÜLLER, Kurt, GABRIEL, Christoph. (ed.) *Multilingual individuals and multilingual societies*. John Benjamins, 2012, p. 241-257.

HÖDER, Steffen. “Constructing diasystems: Grammatical organisation in bilingual groups”. In: ÅFARLI, Tor A.; MÆHLUM, Brit. *The Sociolinguistics of Grammar*. John Benjamins, 2014, p. 137-152.

MACHADO VIEIRA, Marcia dos Santos; BARBOSA, Juliana Bertucci. “Coleções de dados brasileiras para o ensino de Português”. In: MEIRELES, Vanessa; MACHADO VIEIRA, Marcia dos Santos (ed.). *Variação e ensino de português no mundo*. São Paulo: Blucher Open Access, 2022. <https://www.blucher.com.br/variacao-e-ensino-de-portugues-no-mundo>

MACHADO VIEIRA, Marcia dos S.; BARBOSA, Juliana B.; FREITAG, Raquel M. K.; BORGES, Maria Manuel;

MEDEIROS, Ana Lígia S. “Collections of data open to society: linguistic and sociocultural memory and potential for (re)use”. *Cadernos de Linguística*, v. 2, n. 1, p. e607, 24 Jan. 2022. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/607>

MACHADO VIEIRA, Marcia dos Santos; MEIRELES, Vanessa. Digital portal of states of affairs in portuguese and in romance languages to vary and teach. Linguistweets, terceira conferência internacional de linguística no twitter, 2022a. [https://www.linguistweets.org/programa/h/18/](https://www.linguistweets.org/programa/h/18/#linguisTW1945) #linguisTW1945 <https://twitter.com/marciamv2/status/1599852391205306368>

MACHADO VIEIRA, Marcia dos Santos; MEIRELES, Vanessa. “O português no mundo e de todos nós”. In: MACHADO VIEIRA, Marcia dos Santos. *Ensino de Português: Predicar em (con)texto*. São Paulo: Blucher, 2022b, p.201-2017, capítulo 12.

MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz. “Sociolinguística Variacionista e Gramática de Construções: os desafios e as perspectivas de compatibilização”. In: MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz. (org.) *Dimensões e experiências em Sociolinguística*. São Paulo: Blucher, 2019, 314p. <https://www.blucher.com.br/livro/detalhes/dimensoes-e-experiencias-em-sociolinguistica-1575>

MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz. “Variação no modelo construcionista da linguística

funcional-cognitiva”. In: BRESCANCINI, Claudia Regina; MONARETTO, Valéria Neto de Oliveira. (org.) *Sociolinguística no Brasil: textos selecionados*. Porto Alegre: EDIPUCRS, 2020, p.265-304, capítulo 12.

OLIVEIRA JR, Rogério. Português ou brasileiro - qual é o nome da nossa língua? *Roseta*, v. 5, n.1, 2022. <https://www.roseta.org.br/2022/05/11/portugues-ou-brasileiro-qual-e-o-nome-da-no-ssa-lingua/>

ROBIN, Agnès. *Droit des données de la recherche: Science ouverte, innovation, données publiques*. Larcier Eds, 2022.

SIMÕES, Nelson; STANTON, Michael; RIBEIRO, José Luiz Ribeiro; GRIZENDI, Eduardo; MACHADO, Iara; CIUFFO, Leandro. *A Ciberinfraestrutura no Brasil - a contribuição da RNP*. RNP (<https://www.rnp.br/>), [s.d.]. https://www.rnp.br/arquivos/ciberinfraestrutura_rnp.pdf, <https://www.rnp.br/en/node/1453>

WEINREICH, Uriel. “Is a structural dialectology possible?”. *Word* 10: 388-400, 1954. DOI: 10.1080/00437956.1954.11659535

WIEDEMER, Marcos Luiz; MACHADO VIEIRA, Marcia dos Santos. “Paradigma discursivo como (proto)construção: alternância linguística via práticas sociocomunicativas”. In: MACHADO VIEIRA, Marcia dos Santos; MEIRELES, Vanessa (ed.). *Variação em português e em outras Línguas Românicas / Variation en portugais et dans d'autres langues romanes*. São Paulo: Blucher, 2022, p.233-268, cap. 13. <https://www.blucher.com.br/variacao-em-portugues-e-em-outras-linguas-romanicas>

1 <https://variar.wixsite.com/variar>. Acesso em: 03/01/2023.

2 Ministério da Ciência, Tecnologia e Inovações / Fundo Nacional de Desenvolvimento Científico e Tecnológico <https://antigo.mctic.gov.br/mctic/opencms/fundos/fndct/index.html> Acesso em: 03/01/2023.

3 Termo (do tipo guarda-chuva) de análise gramatical usado (entre outros, pelo funcionalista holandês Simon Dik), para referir um evento ou uma situação. Os estados de coisas podem ser classificados como ações, processos, posições, estados, a depender de parâmetros como controle do participante-sujeito e dinamicidade, por exemplo.

4 Aproveitamos o ensejo para agradecer as colaborações de toda a equipe do projeto VariaR (cf. <https://variar.wixsite.com/variar/projetos>) e, especialmente, as dos pesquisadores (docentes e discentes) que se debruçam, solidariamente, sobre questões relativas ao aparato tecnológico a vir dar realidade ao portal digital *inCorpora*: Diogo Vieira (CEFET-RJ – Informática); Erick Gaiote dos Santos (UFRJ – Ciência da Computação); Maria Luiza Machado Campos (UFRJ – Ciência da Computação; GO FAIR Brasil); Pedro Pop-

polino (UFRJ – Ciência da Computação); Renato Campos Mauro (CEFET-RJ – Informática).

5 <https://www.go-fair.org/fair-principles/> . Acesso em 03/01/2023.

6 <https://www.linguistweets.org/programa/h/18/> ; <https://twitter.com/marciamv2/status/1599852391205306368> . Acesso em 03/01/2023.

7 <https://ufrj.br/> . Acesso em 03/01/2023.

8 <http://www.cefet-rj.br/> . Acesso em: 03/01/2023.

9 “Adaptado de Cyberinfrastructure for Environmental Research and Education. Obtido em: http://www.cyrdas.org/related.documents/reports/cyber_report_new.pdf <http://www.ncar.ucar.edu/cyber/cyberreport.pdf> (NSF)” (<https://www.rnp.br/en/node/1453>)

Português

Exploramos movimentos, trilhas e horizontes relativos ao objetivo do projeto VariaR – Variação em línguas românicas/UPVM-UFRJ – de incorporar, numa rede de ações e interações, sujeitos (pesquisadores, professores e outros interessados em dados (socio)linguísticos), *corpora* representativos de usos dessas línguas e informações sobre as gramáticas delas, sobre os textos em que essas gramáticas emergem e sobre as comunidades que delas se servem para se expressarem diversamente. A meta com essa rede é viabilizar a comparação das línguas românicas em uso no mundo, o amplo conhecimento de seus diassistemas e o trabalho com eles em espaços de ciência e educação respaldado por dados da realidade heterogênea e pluricultural e por textos que espelham essa realidade. Tratamos, particularmente, de ações planejadas para a construção do portal *inCorpora*, que busca abrigar contribuições a permitirem um rico panorama das variações em línguas românicas. Entre essas contribuições, atemo-nos, sobretudo, ao *Portal digital de estados de coisas em português e línguas românicas a variar e ensinar*, um subprojeto que conta com apoio financeiro brasileiro (da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro/Cientista do Nosso Estado 32/2021 e Conselho Nacional de Desenvolvimento Científico e Tecnológico/MCTI/FNDCT 18/2021).

English

We explore movements, routes and visions related to the objective of the VariaR project – Variation in Romance Languages/UPVM-UFRJ – to bring together, in a network of actions and interactions, people (researchers, professors/teachers and others interested in (socio)linguistic data), *corpora* representative of the uses of these languages and information about their grammars, about the texts which demonstrate these grammars, and about the communities that use the languages in question to express themselves

in various ways. The goal of such a network is to enable the comparison of Romance languages in use around the world, the broad knowledge of their diasystems and the work involving them in science and education spaces, supported by data from a heterogeneous and pluricultural reality and from texts that mirror this reality. We deal, in particular, with work planned with a view to the construction of the *inCorpora* portal, which seeks to house contributions forming a rich panorama of variations in Romance languages. Among these contributions, we focus particularly on the Digital portal of States of Affairs in Portuguese and Romance Languages to Vary and to Teach, a subproject with Brazilian financial support (from the *Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro/Cientista do Nosso Estado* 32/2021 and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico/MCTI/FNDCT* 18/2021).

Français

Nous explorons les mouvements, les voies et les horizons liés à l'objectif du projet VariaR – Variation dans les langues romanes/UPVM-UFRJ – d'incorporer, dans un réseau d'actions et d'interactions, des sujets (chercheurs, enseignants et autres personnes intéressées par les données (socio)linguistiques), des corpus représentatifs d'usages de ces langues et des informations sur leurs grammaires, sur les textes dans lesquels ces grammaires émergent et sur les communautés qui les utilisent pour s'exprimer de manière diversifiée. L'objectif de ce réseau est de permettre la comparaison des langues romanes en usage dans le monde, la connaissance approfondie de leurs diasystèmes et le travail avec elles dans les espaces scientifiques et éducatifs, en s'appuyant sur des données issues de la réalité hétérogène et pluriculturelle et sur des textes qui reflètent cette réalité. En particulier, nous traitons des actions prévues pour la construction du portail *inCorpora*, qui vise à abriter des contributions permettant un riche panorama des variations des langues romanes. Parmi ces contributions, nous nous intéressons particulièrement au *Portail numérique d'états de choses en portugais et en langues romanes en variation et en enseignement*, un sous-projet subventionné par un organisme Brésilien (de la Fondation pour le soutien à la recherche de l'État de Rio de Janeiro/*Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro/Cientista do Nosso Estado* 32/2021 et du Conseil national pour le développement scientifique et technologique/*Conselho Nacional de Desenvolvimento Científico e Tecnológico/MCTI/FNDCT* 18/2021).

Mots-clés

variation linguistique, projet VariaR, langues romanes, portugais, portail numérique

Keywords

linguistic variation, VariaR project, romance languages, portuguese, digital portal

Palavras chaves

variação linguística, projeto VariaR, línguas românicas, português, portal digital

Marcia dos Santos MACHADO VIEIRA

Universidade Federal do Rio de Janeiro/FAPERJ e CNPqmarcia@letras.ufrj.br
marcia@letras.ufrj.br

Vanessa MEIRELES

Universidade Paul Valéry – Montpellier 3vanessa.meireles@univ-montp3.fr