

PROBLEMAS, SOLUÇÕES E HIPÓTESES NO ESTUDO DA LINGUAGEM DOS NIPO-BRASILEIROS DO DISTRITO FEDERAL DO BRASIL

PROBLEMS, SOLUTIONS, AND HYPOTHESES IN THE STUDY OF THE LANGUAGE OF JAPANESE-BRAZILIANS FROM THE FEDERAL DISTRICT OF BRAZIL

*Fernando Brissos*¹

Resumo: Apesar de formarem uma das maiores comunidades de origem estrangeira no Brasil, os nipo-brasileiros não têm merecido o previsível destaque nos estudos de linguística variacional. Uma exceção honrosa são os trabalhos de Yuko Takano, tanto na sua tese de doutoramento (2013), que procede a um *Esboço do Atlas do Falar dos Nipo-Brasileiros do Distrito Federal*, como no seu pós-doutoramento (2018), em que o amplo *corpus* dessa tese é rentabilizado sob a forma de um estudo dialetométrico (ou seja, quantitativo). Ambos os trabalhos são totalmente inovadores e não têm até hoje qualquer paralelo: não existe ainda outro atlas linguístico nem outro estudo dialetométrico do sistema linguístico nipo-brasileiro. Neste artigo tomamos o *corpus* de Takano e discutimos os seus *achievements* metodológicos e analíticos. Comparamos o primeiro conjunto de *achievements* com os conhecimentos que século e meio de dialetologia científica nos legaram, discutindo os ganhos e perdas de eficiência originados pelas adaptações metodológicas efetuadas por Takano com vistas ao estudo de um sistema linguístico tão específico. No que respeita ao segundo grupo de *achievements*, ou seja, aos resultados da análise quantitativa, expandimos o *corpus* utilizado por Takano de modo a incluir os sistemas linguísticos japonês e português – i.e. os sistemas matriz da variedade mista nipo-brasileira –, tornando assim possível testar os resultados do estudo dialetométrico da autora à luz da genética da variedade nipo-brasileira, por um lado, e, por outro lado, comparar os padrões geolinguísticos da variedade com os das outras variedades brasileiras que foram já sujeitas ao mesmo tipo de análise (Região Sul do Brasil e estado do Amazonas). Desse processo resulta um argumento fundamental: é necessário criar tradição de estudo do sistema linguístico nipo-brasileiro, que está sob forte pressão e tem um interesse cultural de exceção.

Palavras-chave: Dialetometria. Dialetologia brasileira. Variedades linguísticas nipo-brasileiras. Distrito Federal do Brasil.

Abstract: Despite accounting for one of the largest communities of foreign origin in Brazil, Japanese-Brazilians have not yet been put under the spotlight by dialectological studies. The

1 Docente da Faculdade de Letras da Universidade de Lisboa. Endereço de e-mail: fernandobrissos@campus.ul.pt ORCID iD: <https://orcid.org/0000-0002-2525-1987>

most notable exception to that underperforming of Japanese-Brazilian dialectology is Yuko Takano's work. The author has recently (2013) published a PhD thesis under the title of *Esboço do Atlas do Falar dos Nipo-Brasileiros do Distrito Federal* (= *Outline of the Atlas of the Language of Japanese-Brazilians from the Federal District*), whose corpus was subject to due dialectometrical – i.e., quantitative – processing in her Post-Doc dissertation (2018). Both studies remain unparalleled: there is no other linguistic atlas or dialectometric study of the Japanese-Brazilian linguistic system yet. In this paper we take on Takano's corpus and discuss its methodological and analytical achievements. We compare the first set of achievements with the knowledge that a century and a half of scientific dialectology bequeathed to us, discussing the gain and loss in efficiency brought about by the methodological adaptations that were undertaken by Takano in order to study such a specific linguistic system. With regard to the second group of achievements, i.e., the results provided by Takano's analyses, we expand the author's corpus to include the Japanese and Portuguese language systems – that is, the base-systems of the Japanese-Brazilian mixed variety –, thus making it possible to (i) test the results of the author's dialectometric study in light of the genetics of the Japanese-Brazilian variety and (ii) compare the dialectological patterns of that variety with the ones found in the other Brazilian varieties that were subjected to the same type of analysis (Southern Brazil and the Amazonas state). We therefore proceed to a fundamental argument: it is necessary to quickly develop a strong body of Japanese-Brazilian dialectological studies, as the respective linguistic varieties are under strong pressure and have an exceptional cultural interest.

Keywords: Dialectometry. Brazilian dialectology. Japanese-Brazilian linguistic varieties. Federal District of Brazil.

1. INTRODUÇÃO

É falada em várias regiões do Brasil, pelos imigrantes japoneses e seus descendentes (a comunidade *nikkei*), uma variedade linguística mista originada pelo contacto entre a língua japonesa, dialetal (*hogen*) e *standard*, e a portuguesa, também dialetal e *standard*. Do ponto de vista formal, essa variedade caracteriza-se pelo cruzamento da gramática japonesa (o substrato) com o léxico português (superstrato). Chamado tradicionalmente de *koroniago* (“língua da colónia”) mas atualmente, devido ao possível preconceito negativo dessa expressão, também de “falar nipo-brasileiro” ou “variedade nipo-brasileira” (TAKANO, 2018, p. 16, n. 10), o sistema linguístico em questão tem um evidente interesse linguístico, pois resulta da interação de dois mundos muito distantes. Tem ainda, porém, pouca tradição de estudo. São exceção importante os trabalhos de Yuko Takano, sobretudo o *Esboço do Atlas do Falar dos Nipo-brasileiros do Distrito Federal: Aspecto Semântico-lexical* (TAKANO, 2013), que é o primeiro e único estudo de natureza perfeitamente geolinguística sobre a variedade nipo-brasileira, e o relatório de pós-doutoramento, ainda inédito, em que a autora analisa quantitativamente, à luz da dialetometria, os dados do atlas referido (TAKANO, 2018, trabalho efetuado na Universidade de Lisboa sob a supervisão deste autor e de João Saramago). Trata-se de trabalhos de rutura, que lidam com um conjunto de questões novas, quer no que toca à metodologia de recolha de

dados geolinguísticos, quer no respeitante à análise desses dados. No presente artigo ocupar-nos-emos precisamente daqueles trabalhos, que analisaremos nos dois âmbitos referidos (metodologia de recolha de dados e análise de dados), comparando os desafios colocados pelo tratamento geolinguístico da variedade nipo-brasileira com os procedimentos já bem estabelecidos do tratamento geolinguístico de comunidades monolíngues como, por exemplo, a maioria da população brasileira. Compararemos, portanto, o que sabemos e o que precisamos de fazer para saber mais sobre a variedade nipo-brasileira, por um lado, com o que sabemos e o que é preciso fazer para saber mais sobre o contexto geolinguístico em que ela se insere – o sistema dialetal brasileiro –, por outro.

Referiremos previamente, no entanto, cinco pontos importantes e relativamente consensuais sobre a história da imigração japonesa no Brasil, tomando como referência principal a extensa e cuidada introdução de Takano (2013, p. 27-67):²

- a) a imigração japonesa para o Brasil deu-se sobretudo no século XX, tendo sido organizada centralmente pelos governos de ambos os países num processo sistemático que decorreu entre 1908 e 1973. Da mesma forma que a economia brasileira necessitava de mão-de-obra estrangeira no virar do século XIX, veio a ter excesso de mão-de-obra no último quarto do século XX, encontrando-se a economia japonesa no total oposto: excesso de mão-de-obra no final do séc. XIX, depois da abertura do país ao exterior na sequência da revolução Meiji de 1868 e do desenvolvimento económico subsequente, e necessidade de importação de trabalhadores no final do séc. XX, como qualquer outra economia de alto rendimento. Deu-se portanto, a partir da década de 1980, uma inversão do fluxo migratório entre Japão e Brasil: é agora este que se encontra na posição de exportador de trabalhadores para o primeiro;
- b) a comunidade nipo-brasileira (imigrantes japoneses e descendentes no Brasil) é a maior comunidade de origem nipónica fora do Japão, situação que se verificava já antes da 2.^a Guerra Mundial; vivem no Brasil, presentemente, mais de 1.500,000 nipo-brasileiros (MIURA; CATARINO, 2010, p. 66 *apud* TAKANO, 2013, p. 46), dos quais a maior parte se encontra no estado de São Paulo (sobretudo) e nos estados próximos do Paraná e do Mato Grosso do Sul. Em Brasília, vivem desde a fundação da cidade (1957);
- c) o motivo da migração dos nipo-brasileiros para Brasília, provenientes na sua maioria de São Paulo, foi o mesmo da sua imigração para o Brasil: a agroindústria, setor económico em que o Japão tinha performance de ponta já no final do séc. XIX. A comunidade nipo-brasileira de Brasília é a maior comunidade de origem estrangeira presente no Distrito Federal (DF), com c. 0,4% da população = c. 10.000

2 Podem ver-se mais detalhes sobre a história da imigração japonesa no Brasil em Brissos (2021b) ou no passo citado de Takano (2013), com bibliografia.

habitantes no início do século (JOKO-VELTMAN; TORREÃO; SUGIMOTO, 2008, p. 103; KUYAMA, 2000, p. 11 *apud* TAKANO, 2013, p. 53);

- d) existem dois principais blocos temporais na imigração japonesa para o Brasil: o antes e o pós-2.^a Guerra Mundial, quando surgem os chamados “imigrantes permanentes” (*ijyuu-imin*). Até à 2.^a Guerra Mundial, o objetivo fundamental dos imigrantes japoneses no Brasil era trabalhar, acumular reservas financeiras e voltar para o país de origem; depois da guerra, com a crise económica que afetou o Japão (país que ficou do lado derrotado), esses imigrantes interiorizam a improbabilidade de voltar a residir na terra natal. Dá-se então uma progressiva assimilação dos nipo-brasileiros no segundo país da equação, o que se traduziu, desde logo, num êxodo rural; agora era fundamental falar a língua portuguesa, escolarizar os filhos no sistema educativo brasileiro, obter empregos qualificados no terceiro setor, ocupar um lugar de pleno direito na sociedade brasileira. Mudou então o perfil do nipo-brasileiro, que se tornou urbano, com alta escolarização e muito mais miscigenado; e as comunidades de imigrantes japoneses deixaram de ser uma extensão do Japão para passar a ser mais uma cor progressivamente diluída no complexo mosaico cultural brasileiro;
- e) com essa assimilação tem corrido paralelo um progressivo abandono do japonês como língua veicular da comunidade nipo-brasileira, que, mesmo quando conservado, sofre cada vez mais influência do português. Um processo plenamente previsível, pois a língua é, acima de tudo, uma ferramenta de otimização dos recursos sociais, desde logo a comunicação; se não tem valor acrescentado num determinado meio social, uma língua tende, tal como qualquer outro *asset*, a ser eliminada ou, pelo menos, a ver o seu raio de ação crucialmente reduzido.

A aventura japonesa no Brasil tem tido portanto, como todas as aventuras, consequências linguísticas, que o linguista egoistamente agradece; a oportunidade de estudar a colisão entre dois sistemas linguísticos tão diferentes é interessante e não pode ser enfeitada, apesar dos importantes desafios colocados. Os primeiros desafios são os que dizem respeito à metodologia de recolha dos dados, que passamos a ver.

2. QUESTÕES METODOLÓGICAS NA RECOLHA DE DADOS GEOLINGUÍSTICOS

A dialetologia é uma ciência mais elegante e elástica do que tantas vezes se pensa, e o atlas de Takano (2013) faz-lhe jus. A autora deparou-se com dois problemas complexos no que respeita à metodologia de recolha de dados geolinguísticos, que resolveu de forma simples e, acima de tudo, eficaz; a robustez dos resultados que podemos extrair do *corpus* de dados do atlas, de que trataremos na secção 3, demonstra-o.

O primeiro problema reside no instrumento fundamental da recolha dos dados de qualquer atlas linguístico, ou seja, o questionário. Subjacente a esse problema está, no caso de Takano (2013), um par de fatores decisivos: a coexistência de dois sistemas linguísticos que não são mutuamente inteligíveis (o japonês e o português) e a progressiva substituição de um por outro (i.e. a sobreposição do português ao japonês). Haveria então que formular um questionário que não potenciasse o uso de uma ou outra língua e, mais do que isso, fomentasse o uso da variedade em estudo. Assim, a autora formulou um “questionário semântico-lexical visual” (QSLV), ou seja, um inventário de questões linguísticas formuladas por meio de figuras e não por elicitación linguística, como sucede na generalidade dos atlas linguísticos. Pode ver-se um exemplo do questionário de Takano na figura 1; todas as figuras utilizadas pela autora foram retiradas de materiais didáticos para o ensino de japonês a estrangeiros (cf. TAKANO, 2013, p. 36 e 109).



Figura 1: Exemplo de imagem utilizada no questionário de Takano (2013)
Tema: festa de aniversário. (Fonte: TAKANO, 2013.)

O QSLV contém 219 figuras = 250 questões sobre temas quotidianos, que foram mostradas sucessivamente aos informantes, levando-os a falar (i.e. a *dissertar*) sobre cada tema.³ Assim foi possível obter dados respeitantes não apenas ao funcionamento do sistema linguístico utilizado em cada resposta (raiz lexical, afixos, fonologia, etc.), mas também ao próprio uso de cada sistema linguístico, i.e. à escolha que os informantes tendem a fazer, no discurso normal e não-especializado do dia-a-dia, entre o uso do japonês, do português ou do sistema misto (a variedade nipo-brasileira). A utilização de figuras com temas quotidianos tem outra virtude: possibilitar conversa aberta e espontânea, aspeto nem sempre devidamente acautelado na recolha de *corpora* de variação

3 A lista dos temas é especificamente: a) convivência: vida quotidiana (a-1, domínio domiciliar; a-2, domínio social); b) convivência: marcas culturais (b-1, domínio domiciliar; b-2, domínio social); c) utensílios do quotidiano/materiais. Vejam-se mais detalhes em Takano (2013, esp. p. 109-110).

linguística. Com efeito, a mera resposta automática ou semiautomática a uma lista de questões não proporciona o mesmo grau de fiabilidade linguística que a produção de discurso tão espontâneo quanto possível, pois é espontâneo o uso normal que fazemos da língua no dia-a-dia. Percebe-se, então, que Takano (2013) não tenha utilizado um questionário especializado, com respostas tendencialmente focadas e curtas, como é tradicional na geolinguística: a grande questão no estudo da variedade nipo-brasileira é ainda, mais do que uma descrição gramatical exaustiva, identificar onde a mesma é falada e por quem é falada (sexo, idade, etc.); depois de feita essa catalogação é que se pode descer à minúcia daquela descrição. Eis aqui uma razão primordial para que surjam mais atlas ou outros *corpora* de grandes dimensões da variedade nipo-brasileira, almejando cobrir, tanto quanto possível, todo o país.

O principal contra da utilização de um questionário baseado em figuras ou desenhos é a dificuldade acrescida de identificar a resposta dada pelo informante ao conceito específico que é alvo de cada figura. O investigador tem de analisar com cuidado a totalidade das respostas – que por vezes são longas (ou *dissertativas*) ou fogem mesmo ao tema-alvo –, idealmente transcrevê-las e, a partir daí, escolher tantas variantes quantas as utilizadas pelos informantes para cada variável (i.e. todas as respostas diferentes produzidas para cada conceito/tema). A alternativa seria pior, todavia: correr o risco de elicitare o uso de um ou outro sistema linguístico numa comunidade que está sob forte pressão nesse plano. Para além disso, a obtenção de respostas em discurso semidirigido tem tradição na geolinguística; pode tomar-se como exemplo o ALEPG – *Atlas Linguístico-Etnográfico de Portugal e da Galiza*, i.e. o projeto de atlas linguístico nacional de Portugal, que apoia boa parte dos seus resultados (concretamente os materiais obtidos em c. 2/3 dos inquéritos) precisamente nesse tipo de discurso (pode ver-se uma descrição detalhada do ALEPG em SARAMAGO, 2006).

Tem menos tradição a solução encontrada por Takano (2013) para o segundo problema de especial complexidade que encontrou na constituição do *corpus*: a obtenção de dados em informantes de ambos os sexos. Tanto o problema como a solução são, não obstante, fáceis de explicar: por razões que se prendem com tradições culturais da sociedade nipónica, a autora deparou-se com uma dificuldade sistemática na obtenção de informantes do sexo masculino, que frequentemente recusavam a sua participação no projeto; ora, entre ter e não ter dados de todo em tempo útil, a autora optou por constituir o *corpus* a partir apenas do sexo feminino. Uma vez que, tradicionalmente, predominam os informantes do sexo masculino nos atlas linguísticos, o perfil de informante de Takano (2013) limita, para todos os efeitos, a comparabilidade dos dados; mas, como dissemos no início desta secção, a dialetologia é uma ciência suficientemente maleável para, perante a escolha entre ter alguns dados ou não ter nenhuns dados, possibilitar a única opção que faz sentido: ter dados. Estudos posteriores alargarão, certamente, o campo de recolha e incluirão informantes masculinos, tal como os estudos geolinguísticos do português e de qualquer outro sistema linguístico também têm cada vez mais

de incluir informantes femininas; é hoje claro que só com ambos os sexos representados um *corpus* linguístico consegue ter a totalidade do objeto de estudo representado.

As restantes opções metodológicas previamente tomadas por Takano (2013) a respeito do perfil de informante puderam ser mantidas. Os informantes do atlas são, então, mulheres nipo-brasileiras de 2.^a geração (*nissei*) bilíngues que passaram pelo menos metade da sua vida na região-alvo e têm pouca escolaridade em língua japonesa. A escolaridade em língua portuguesa não foi controlada, tal como as profissões. Estes dois aspetos suscitam questões, pois estão em causa variáveis importantes no tocante à competência linguística, mas, tendo em conta que os campos semânticos alvo do questionário dizem todos respeito à vida quotidiana, o problema é minimizado. Em todo o caso, espera-se que estudos posteriores controlem cada vez mais variáveis, de modo a termos um panorama tão exato quanto possível dos sistemas linguísticos utilizados pelos nipo-brasileiros no Brasil.

Takano (2013) utiliza um total de 10 informantes = 2 por localidade, divididos em duas faixas etárias: 51-65 anos e mais de 65 anos. As cinco localidades do atlas, que são, na verdade, cinco regiões, devido à dispersão das comunidades nipo-brasileiras nas zonas periféricas de Brasília (uma vez que o seu destino original foi, como vimos, o trabalho no setor agroindustrial), estão indicadas no mapa 1⁴. Três dessas regiões são rurais – Brazlândia, Taguatinga e Vargem Bonita – e duas são urbanas – Brasília e Núcleo Bandeirante (antiga Cidade Livre). Encontramos, então, polivalência no tipo geográfico e socioeconómico das localidades (rural/urbano); um número significativo de pontos de inquérito para uma região pequena como o Distrito Federal (que tem apenas 5.760,784 km², de acordo com os dados oficiais do IBGE); e duas faixas etárias localizadas acima dos 50 anos, o que está de acordo com a metodologia da geolinguística clássica (que utiliza informantes acima do meio da idade de forma a registar variedades mais conservadoras e, por conseguinte, mais próximas do dialeto tradicional de cada região). Por isso, apesar das naturais dificuldades colocadas por um tema verdadeiramente inovador, Takano (2013) deixa-nos uma quantidade assinalável de dados; se não considerarmos a existência de eventuais respostas nulas, 219 questões * 10 informantes = 2.190 respostas indexáveis a 5 localidades do DF. A única forma de conseguirmos digerir um *corpus* desta natureza é a análise quantitativa.

3. TRATAMENTO QUANTITATIVO DO *CORPUS* E ANÁLISE DOS RESULTADOS

O atlas de Takano (2013) é, como o típico atlas linguístico, um *corpus* de grandes dimensões, i.e. aquilo a que poderemos chamar, em linguagem atual, um arquivo de *big data* dialetais; e quaisquer *corpora* de grandes dimensões só têm o devido rendimento se forem sujeitos a processamento quantitativo. A única metodologia de análise

4 Todos os mapas apresentados neste trabalho constam do anexo colocado no final do texto.

quantitativa de dados dialetais suficientemente testada até hoje é a dialetometria, que Brissos e Saramago (2019, p. 353) definem como

uma abordagem quantitativa ao estudo dos dialetos com um enfoque na métrica, i.e. na mensuração dos fenómenos de variação dialetal por meio de procedimentos exatos e totalmente comparáveis, os quais importa da classificação numérica ou taxonómica. Aplica cálculos matemático-estatísticos elaborados à matriz de dados obtida a partir dos procedimentos referidos e representa cartograficamente (*espacializa*) os resultados desses cálculos, cabendo ao linguista, com a liberdade que a estatística confere, a tarefa final de interpretação do quadro geolinguístico que tem à frente.

O ponto essencial é, portanto, a abordagem métrica (ou seja, uniformizada) a materiais linguísticos, que permite a constituição de bases de dados que, por sua vez, podem ser sujeitas a tratamento quantitativo. No domínio românico, a corrente teórico-metodológica predominante nos estudos dialetométricos tem sido a chamada EDS – *Escola de Dialetometria de Salzburgo* (GOEBL, 1984, 2016), que, no caso português, é mesmo a única com efetiva tradição (e.g. BRISSOS, 2016; BRISSOS; GILLIER; SARAMAGO, 2017; BRISSOS; SARAMAGO, 2019). Takano (2018) insere-se nessa corrente e apresenta uma digestão dialetométrica dos materiais de Takano (2013) formada a partir de uma base de dados com as seguintes características:

- a) consideração apenas da 1.^a faixa etária (51-65 anos), uma vez que se trata de um estudo exploratório e é, por isso, conveniente conter o número de variáveis analisadas num mínimo funcional;
- b) consideração de todos os pontos de inquérito utilizados;
- c) consideração apenas de dados lexicais, i.e. de variação respeitante à raiz dos vocábulos utilizados nas respostas ao questionário;
- d) 97 questões consideradas, após excluídas as que não obtiveram resposta em todos os pontos de inquérito (e, por isso, poderiam distorcer a análise estatística) e as que obtiveram a mesma resposta em todos os pontos (não permitindo, portanto, identificar áreas linguísticas diferentes);
- e) as respostas foram subdivididas em três variantes: resposta em japonês; resposta em português; resposta mista (variedade nipo-brasileira);⁵
- f) foram excluídas as respostas múltiplas (i.e. a existência de mais de uma variante para cada variável), utilizando-se apenas a 1.^a resposta a cada questão.

5 Exemplo para o tema 15 “Cortar o papel” do questionário: *cortando* (registado em Brazlândia) = sistema português; *kiru* (Núcleo Bandeirante, Taguatinga e Vargem Bonita) = sistema japonês; *corta shitoru* (Brasília) = sistema misto (TAKANO, 2018, p. 31).

Apenas o aspeto mencionado em e) merece, à luz do paradigma de Salzburgo, atenção detalhada, pois não deixa de levar à omissão de interessante variação dialetal, quer no caso do japonês, do português ou mesmo, possivelmente, do sistema misto. Com efeito, um agrupamento das respostas em apenas três macrogrupos oculta as eventuais microvariações que possam ser encontradas nos dados: será mais do que natural que um informante originário de São Paulo produza, para uma mesma questão, uma resposta diferente de um informante originário do Mato Grosso, mesmo que ambos vivam no DF há décadas e estejam a utilizar o mesmo sistema linguístico. Duas razões, no entanto, validam o procedimento seguido: por um lado, o facto de o DF ser, para todos os efeitos, uma região de pequena dimensão e, por isso, tipicamente menos sujeita a variação espacial; por outro lado – e mais importante –, o facto de, como vimos na secção 2, a questão fundamental no estado atual do conhecimento sobre a variedade nipo-brasileira ser a identificação das áreas e das condições sociais da sua ocorrência. Nesse sentido, um primeiro passo como o estudo de Takano (2018) é plenamente razoável.

Podemos sintetizar os resultados da autora em dois aspetos fundamentais, um absoluto e um relativo. O primeiro pode ser designado de absoluto porque não é, de todo, imprevisível: Takano (2018) verifica que as relações de parentesco linguístico dos pontos de inquérito utilizados tendem a definir-se pela posição geográfica, i.e. pontos mais próximos geograficamente tendem a ser mais próximos linguisticamente. Trata-se tão-somente da mera verificação de uma lei dialetológica bem conhecida: a distância linguística tende a ser correlacionável da distância física ou geográfica.

O segundo aspeto é mais interessante, pois diz respeito a especificidades do *corpus* em análise: encontra a autora uma cisão fundamental entre a cidade de Brasília e os entornos (i.e. a periferia). O mapa 2, que representa a aplicação ao *corpus* de uma análise dendrográfica com segmentação em 2 agrupamentos, mostra-nos essa divisão: Brasília fica isolada num *cluster* (cor verde) e todo o resto do *corpus* forma o restante *cluster* (cor vermelha).⁶ Takano fornece uma hipótese explicativa para essa cisão que se apoia na demografia de cada região:

[...] a zona da Capital é uma região extremamente urbana e o povoamento da comunidade nipo-brasileira estende-se ao longo dessa região. Os (i)migrantes vieram para exercer profissões diversas e dispersas, enquanto que nas zonas dos Entornos o povoamento dos (i)migrantes ocorreu de forma uniforme e concen-

6 A análise dendrográfica, ou *cluster analysis*, é um dos parâmetros mais utilizados na dialetometria de Salzburgo. Permite segmentar o *corpus* em agrupamentos (*clusters*) de pontos de inquérito mais ou menos homogéneos linguisticamente; quanto maior o número de agrupamentos (ou seja, quanto menor o número de pontos de inquérito por agrupamento), maior homogeneidade em cada agrupamento e vice-versa. Veja-se mais informação sobre este parâmetro dialetométrico em Goebel (2016, p. 85-86) e Brissos, Gillier e Saramago (2017, p. 17).

trada. Esse contexto social pode ter contribuído para a configuração dialetal dos nipo-brasileiros do Distrito Federal. (TAKANO, 2018, p. 35)

Independentemente da natureza histórica ou demográfica de cada nuance dialetal que encontramos na região, é claro que o estudo da variedade nipo-brasileira do Distrito Federal beneficiaria de uma perspectiva comparativa com um panorama mais lato, formado pela relação com os sistemas linguísticos que estão na origem dessa variedade e lhe servem de variedades-teto: o japonês e o português. No presente trabalho tomamos essa perspectiva, expandindo a base de dados de Takano (2018) para incluir aqueles dois sistemas linguísticos. O procedimento foi simples: acrescentamos dois pontos de inquérito artificiais aos dados de Takano (2018), um correspondente ao sistema japonês e, por isso, com resposta “sistema japonês” para todas as 97 questões do *corpus*, e outro correspondente ao sistema português e, dessa forma, apenas com a resposta “sistema português” nas questões do *corpus*. O mapa 3 representa a rede de pontos de inquérito (segmentando, ao contrário de TAKANO 2018, as áreas geográficas em poligonação de Voronoi, a representação *standard* na dialetometria de Salzburgo); passamos a ver os resultados respetivos.

A primeira questão de investigação é, muito claramente, o porquê de Takano 2018 ter encontrado uma cisão entre Brasília e entornos do DF – e os resultados fornecem uma resposta muito clara. Podemos ver que essa cisão se deve ao grau de preservação da língua japonesa no uso quotidiano da língua: nos entornos, i.e. nas regiões menos urbanas e de povoamento mais concentrado, o japonês tem sido mais conservado do que na cidade de Brasília. O mapa 4, que nos dá a análise dendrográfica do *corpus* em 2 *clusters*, é de fácil interpretação: Brasília identifica-se com o sistema português (*cluster* verde) e todos os entornos se identificam com o sistema japonês (*cluster* vermelho). Se subirmos o número de *clusters*, não deixamos de continuar a verificar uma significativa homogeneidade entre os entornos e o sistema japonês: no mapa 5 (3 *clusters*), é o agrupamento composto por Brasília e Português que se subdivide, mantendo-se o *cluster* formado por entornos e Japonês uniformizado; no mapa 6 (4 *clusters*), Vargem Bonita individualiza-se, mas o Japonês continua a pertencer ao *cluster* mais estendido na região, que se segmenta no mapa 7 (5 *clusters*) mas continua a não individualizar o Japonês; nem no mapa 8, que apresenta o número máximo de *clusters* permitido pelo *corpus* (6 = 7 pontos de inquérito constantes da base de dados menos 1, de forma a permitir um agrupamento de pontos), o Japonês fica isolado, pois forma agrupamento com Taguatinga. Este facto é da maior relevância, pois não apenas nos permite dar uma explicação para a cisão encontrada por Takano 2018 (e reiterada no nosso *corpus*) mas, mais do que isso, deixa evidente que a língua japonesa tem um raio de ação muito lato na variedade nipo-brasileira do DF – muito mais lato do que a língua portuguesa.

A variedade nipo-brasileira do DF é, com efeito, mais japonesa do que se poderia supor, como o mapa 9, que representa a distribuição de assimetria do *corpus*, deixa

ver.⁷ De facto, (i) o ponto correspondente ao sistema japonês tem um valor de assimetria abaixo da média (cor verde), (ii) por oposição ao sistema português, que, juntamente com Brasília, é o único ponto com valores de assimetria acima da média, e logo no escalão máximo (cor vermelha). Isso significa que (iii) o ponto japonês tem uma integração linguística no conjunto dos dados que é superior não só ao ponto português (aspeto relevante por si mesmo, dado que estamos localizados em pleno ambiente geolinguístico de língua portuguesa) mas, ainda mais significativamente, também ao ponto de Brasília. A variedade nipo-brasileira do Distrito Federal conserva, portanto, as suas raízes de forma notável.

Duas outras áreas do Brasil, a Região Sul (o conjunto formado pelos estados do Paraná, Santa Catarina e Rio Grande do Sul, que abreviaremos por RSB) e o Amazonas, foram já sujeitas a tratamento dialetométrico do tipo que vimos para a variedade nipo-brasileira do DF (i.e. uma comparação sistemática com variedades padrão externas); Brissos e Saramago (2019) ocupam-se da primeira (a partir dos materiais do ALERS – *Atlas Linguístico-Etnográfico da Região Sul do Brasil*: ALTENHOFEN; VILSON; KLASSMANN, 2011) e Brissos (2021a) da segunda (utilizando os materiais do ALAM – *Atlas Linguístico do Amazonas*: CRUZ, 2004), sempre, como no presente trabalho, com recurso a dados exclusivamente lexicais. Se compararmos os resultados respetivos, verificamos, por um lado, que a RSB e o Amazonas têm perfis totalmente diferentes e, por outro lado, que a variedade nipo-brasileira do DF se assemelha muito nitidamente à primeira. Os mapas 10 e 11 dão-nos os coeficientes de assimetria da RSB e do Amazonas em *corpora* que incluem não apenas os pontos de inquérito dos atlas linguísticos regionais respetivos mas também dois pontos artificiais correspondentes ao português padrão europeu (PPE) e ao português padrão brasileiro (PPB), representados, tal como no estudo da variedade nipo-brasileira do DF, por quadrados colocados na periferia dos mapas.⁸ Podemos constatar que, enquanto os dialetos do Amazonas se opõem em bloco às variedades padrão da língua (que são os únicos pontos com valores de assimetria acima da média, localizados, significativamente, no escalão máximo, i.e. na cor vermelha), a paisagem da RSB não se distingue das variedades referidas. Com efeito, neste *corpus* o PPB está localizado no escalão verde, i.e. no primeiro escalão

7 A distribuição de assimetria (ou *distribuição skewness*) é um dos parâmetros clássicos da EDS. Dá-nos uma síntese relacional do grau de integração linguística de cada ponto de inquérito no conjunto dos dados: pontos com valores de assimetria abaixo da média (i.e. pontos localizados na metade esquerda do histograma e, portanto, representados com cores frias no mapa) têm um nível de integração linguística no conjunto acima da média e, inversamente (ou seja, *simetricamente*), pontos com valores de assimetria acima da média (representados com cores quentes) têm um nível de integração linguística abaixo da média. Quanto mais alto o valor de assimetria (quanto mais quente a cor), mais baixo esse nível, e vice-versa. Para explicações de detalhe sobre o funcionamento deste parâmetro, vejam-se Goebel (2016, p. 84-85) e Brissos, Gillier e Saramago (2017, p. 19).

8 No caso do *corpus* do Amazonas, o PPE corresponde ao quadrado superior e o PPB ao quadrado inferior; no caso da RSB, a localização é inversa: o PPB corresponde ao quadrado superior e o PPE ao quadrado inferior.

abaixo da média de assimetria, e o PPE, mesmo tendo valores de assimetria acima da média do *corpus*, (i) encontra-se integrado no primeiro escalão acima dessa média (cor amarela), que (ii) inclui outras 107 localidades (é o segundo escalão mais numeroso); (iii) mesmo não sendo uma variedade dialetal da RSB, o PPE não está sequer no lote de variedades com grau máximo de distinção no conjunto (os 8 pontos representados com cor vermelha). Ao contrário do Amazonas, mas tal como no DF, na RSB variedades linguísticas externas apresentam boa capacidade de penetração.

Detetamos o mesmo nexos quando comparamos os resultados da análise dendrográfica das três regiões. Os mapas 12 e 13, que nos dão a segmentação em 2 *clusters* da RSB e do Amazonas, mostram, mais uma vez, uma oposição em bloco dos dialetos amazonenses às variedades *standard* e, em sentido contrário, um acentuado *overlapping* dessas variedades com a RSB, tal como verificámos no caso da variedade nipo-brasileira do DF. É claro que os dados da RSB não são totalmente comparáveis com os do Amazonas e do DF, pois contêm muitos mais pontos de inquérito (277 contra 11 do Amazonas e 7 do DF), mas as diferenças de padrões detetáveis são tão nítidas que permitem, pelo menos, colocar a hipótese de que a variedade nipo-brasiliense se assemelha com o sul do Brasil em aspetos importantes do seu padrão de variação geolinguística (menor distinção de elementos externos), opondo-se, portanto, ao norte. Esta questão excede o mero âmbito do estudo da variedade nipo-brasileira (do DF e não só) e está pendente do avanço nos estudos quantitativos/dialetométricos dos dialetos do português brasileiro; mas depende de uma hipótese sustentável e, por isso, deverá ser retomada à medida que mais estudos dialetométricos sobre variedades linguísticas do Brasil, seja de que língua for, vão surgindo.

No caso do Distrito Federal não parece, contudo, que possamos falar de apenas *uma* variedade nipo-brasileira. Tanto a análise dendrográfica como a distribuição de assimetria são claras em mostrar que existem dois blocos principais na variedade nipo-brasiliense: o de Brasília e o dos entornos (como TAKANO, 2018 já havia notado). Esses blocos distinguem-se pela relação que estabelecem com os sistemas matriz: proximidade com o japonês, no caso dos entornos, ou com o português, no caso de Brasília. Apenas dados detalhados, com mais pontos de inquérito e com a inclusão do sexo masculino e de faixas etárias abaixo dos 50 anos, poderão esclarecer definitivamente esta questão; mas, mais uma vez, não deixamos de poder colocar uma hipótese sustentável a partir de dados que pareciam difíceis de analisar. Este é o principal ponto deste artigo, que retomamos seguidamente.

4. OBSERVAÇÕES FINAIS

Apesar das dificuldades inerentes ao estudo de um sistema linguístico como o resultante do contacto entre os imigrantes japoneses no Brasil e os falantes de português do país (um contacto entre dois mundos muito distantes, como dizíamos no início do

texto), fica claro que é possível adotar os métodos geolinguísticos mais exigentes nesse estudo. Os resultados de Takano (2018) e os deste trabalho mostram-no, quer no plano da recolha de materiais – não obstante a dificuldade de coleção de dados em perfis variados de informantes – quer no que toca à extração de conclusões e hipóteses cientificamente válidas a partir desses materiais. Este facto é de grande relevância, pois a metodologia de análise que seguimos, a metodologia quantitativa de análise de dialetos conhecida como dialetometria, foi criada no que poderíamos chamar de ambiente geolinguístico *ideal*. Concretamente, a dialetometria foi desenhada como ferramenta de análise dos atlas linguísticos europeus clássicos, com os seus questionários relativamente lineares e os seus informantes de tipo homogéneo (o chamado perfil NORM = non-mobile, older, rural male). A transferência da dialetometria para *corpora* menos sistemáticos como os resultantes do contacto de línguas não tem ainda tradição significativa, mas Takano (2018) e este trabalho provam a eficiência da disciplina para lá do seu meio; isso é extremamente útil, pois a única forma de digerir eficientemente *corpora* de grandes dimensões (no caso, *big data* dialetais) como os atlas linguísticos é o recurso a análise quantitativa, e a única metodologia de análise quantitativa de dialetos que tem provado consistentemente a sua robustez são os princípios métricos. Do contacto entre japonês e português resulta, portanto, o contacto entre dialetometria e *corpora* não tradicionais, que poderíamos designar por *corpora de compromisso*; quer dizer, *corpora* assumidamente não ideais mas sim reais, i.e. conjuntos de dados que resultam de um compromisso estabelecido pelo investigador em prol da recolha dos materiais possíveis em vez de materiais ideais que poderiam ser inatingíveis e, por isso, nulos. A dialetologia é uma ciência muito pragmática; e a dialetometria também tem de o ser, pois é uma mera ferramenta da dialetologia. Do contacto entre dialetometria e *corpora* ‘de compromisso’ ganham então ambas as partes, o que é crucial.

No que respeita aos resultados linguísticos concretos do presente estudo, ressaltaremos cinco pontos:

- a) em primeiro lugar, a verificação de que a variedade nipo-brasileira do Distrito Federal do Brasil não é um todo homogéneo, segmentando-se em dois grupos basilares: a cidade de Brasília e os *entornos*,
- b) segmentação essa que se deve ao grau de identificação das subvariedades faladas nas diferentes localidades com os sistemas matriz: enquanto os entornos são identificáveis com o japonês, Brasília é agrupável com o português.
- c) A segmentação é tão nítida, ou seja, depende de um contraste tão forte, que não parece ser possível falar de apenas *uma* variedade nipo-brasileira mas, aparentemente, de *duas*.
- d) Em qualquer caso, ambos os sistemas matriz têm uma boa capacidade de penetração na região – que, por isso mesmo, é vincadamente plural.
- e) Essa pluralidade e essa abertura a variedades externas tornam o panorama geolinguístico do sistema nipo-brasileiro do DF parecido com o da Região Sul do

Brasil e diferente do do maior estado do norte (o Amazonas), i.e. as duas regiões brasileiras que foram já sujeitas a tratamento dialetométrico comparável.

O terceiro ponto forma o aspeto mais relevante no que respeita à caracterização do sistema linguístico nipo-brasiliense propriamente dito e deverá ser verificado por estudos que completem os dados de Takano (2013); mas já vimos que esses estudos podem ser feitos com mais elasticidade do que tradicionalmente se pensa, pois a digestão quantitativa dos materiais respetivos é realizável sem falta de eficiência.

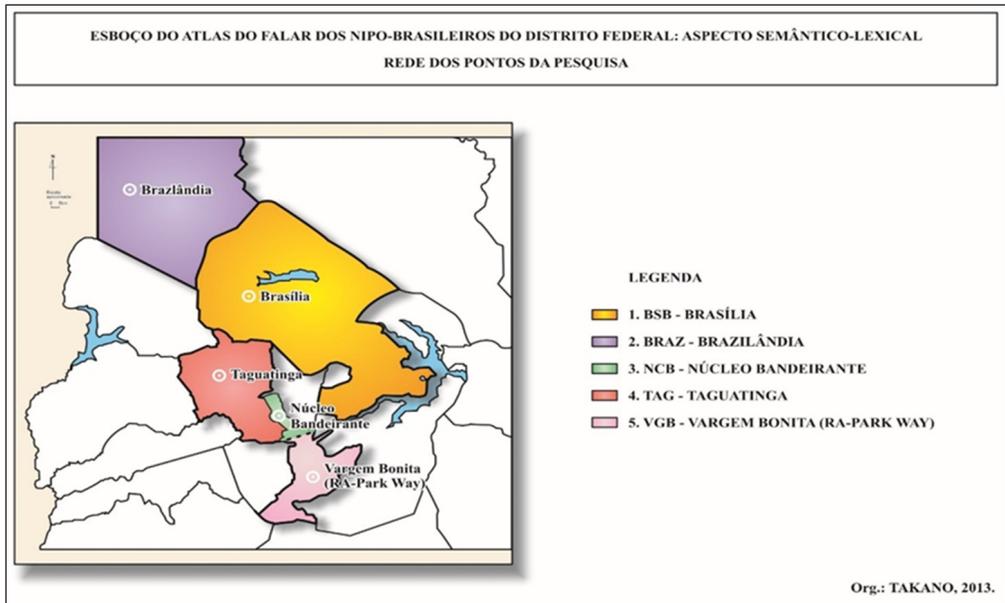
Por isso a espantosa aventura japonesa no Brasil está, no que respeita ao linguista, apenas a começar; e não terminará quando finalmente se concretizar o tão desejado *atlas linguístico das variedades nipo-brasileiras*, tarefa que urge começar. Em sentido mais geral, é imperioso começar também a dotar o sistema linguístico nipo-brasileiro de um corpo de estudo significativo, pois a aventura japonesa é, pela sua história tão idiossincrática, uma preciosidade não apenas cultural mas também científica; e a progressiva assimilação dos nipo-brasileiros na sociedade e na cultura do Brasil tende a eliminar estratos linguísticos que serão do maior interesse para o estudo do contacto de línguas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALTENHOFEN, C. V.; KLASSMANN, M. S. (Orgs.). **Atlas Linguístico-Etnográfico da Região Sul do Brasil (ALERS)**: Cartas semântico-lexicais. UFRGS Editora / Editora da USFC, 2011.
- BRISSOS, F. Portugal: a cidade e o interior. I – Centro-sul. **Limite**, Universidad de Extremadura, n. 10.1, p. 85-107, 2016. Disponível em: <<http://www.revistalimite.es/volumen%2010/05brissos.pdf>>.
- BRISSOS, F.; GILLIER, R.; SARAMAGO, J. Variação lexical açoriana: estudo dialetométrico do Atlas Linguístico-Etnográfico dos Açores. In: ÁLVAREZ DE LA GRANJA, M.; BOULLÓN AGRELO, A.; GONZÁLEZ SEOANE, E. (Eds.). **Aproximacións á variación lexical no dominio galego-portugués**. A Coruña: Área de Filoloxías Galega e Portuguesa / Departamento de Galego-Portugués, Francés e Lingüística / Universidade da Coruña. Monografía 11 da **Revista Galega de Filoloxía**, p. 11-28, 2017. Disponível em: <<http://illa.udc.es/rgf/monografias.html>>.
- BRISSOS, F.; SARAMAGO, J. Análise dialetométrica do Atlas Linguístico-Etnográfico da Região Sul do Brasil: variação lexical. In: CARRILHO, E.; MARTINS, A. M.; PEREIRA, S.; SILVESTRE, J. P. (Orgs.). **Estudos linguísticos e filológicos oferecidos a Ivo Castro**. Lisboa: Centro de Linguística da Universidade de Lisboa, 2019, p. 349-379. Disponível em: <<http://hdl.handle.net/10451/39619>>.
- BRISSOS, F. Análise dialetométrica do Atlas Linguístico do Amazonas. **Revista Internacional de Lingüística Iberoamericana**, n. 37, 2021a. No prelo.
- BRISSOS, F. Contando a linguagem dos nipo-brasileiros do Distrito Federal na soma dos dialetos brasileiros. Palestra apresentada ao XIII Congresso Internacional de Estudos

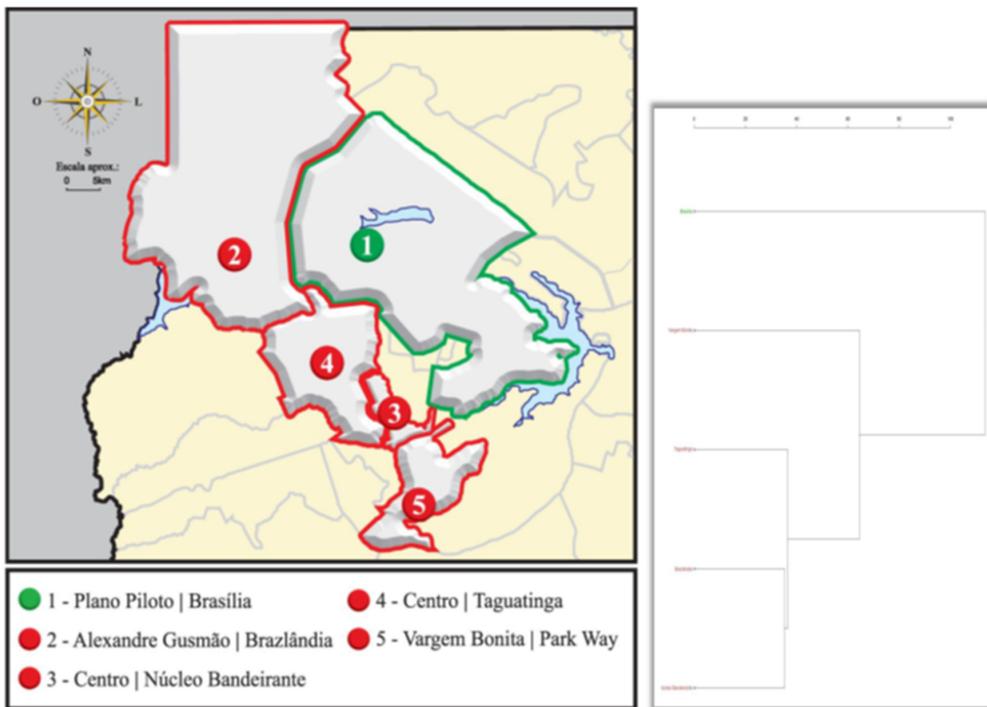
- Japoneses do Brasil / XXVI Encontro Nacional de Professores Universitários de Língua, Literatura e Cultura Japonesa, 17-19/03/2021, Universidade de Brasília. [Transcrição publicada nas **Atas** do congresso, 2021b.] No prelo.
- CRUZ, M. L. de C. **Atlas Lingüístico do Amazonas**. Dissertação de doutoramento inédita, Universidade Federal do Rio de Janeiro, 2004. 2 vols.
- GOEBL, H. **Dialektometrische Studien**: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF. Tübingen: Niemeyer, 1984. 3 vols.
- GOEBL, H. Romance linguistic geography and dialectometry. In: LEDGEWAY, A.; MAIDEN, M. (Eds.). **The Oxford guide to the Romance languages**. Oxford University Press, p. 73-87.
- IBGE Cidades**. In: INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. 2017. Disponível em: <<https://www.ibge.gov.br>>.
- JOKO-VELTMAN, I.; TORREÃO, A.; SUGIMOTO, F. M. 1957: Ano 1 da imigração japonesa no Distrito Federal. In: **Centenário da imigração japonesa no Brasil – Cinquentenário da presença nikkey em Brasília**. Brasília: Thesaurus, 2008.
- KUYAMA, M. **O uso da língua japonesa na comunidade nipo-brasileira**: o empréstimo lexical no japonês falado pelos imigrantes. Dissertação de doutoramento inédita, Universidade de São Paulo, 2000.
- MIURA, H.; CATARINO FILHO, M. R. **Japão e Brasília**: Imigração e esporte. Brasília: Thesaurus, 2010.
- SARAMAGO, J. O Atlas Lingüístico-Etnográfico de Portugal e da Galiza (ALEPG). **Estudis Romànics**, n. 28, p. 281-298. Disponível em: <<https://publicacions.iec.cat/repository/pdf/00000018/00000002.pdf>>.
- TAKANO, Y. **Esboço do Atlas do Falar dos Nipo-Brasileiros do Distrito Federal**: aspecto semântico-lexical. Dissertação de doutoramento inédita, Universidade de São Paulo, 2013. 3 tomos.
- TAKANO, Y. **Atlas Lingüístico Semântico-Lexical do Falar Nipo-Brasileiro do Distrito Federal**: Perspectiva dialetométrica. Dissertação de pós-doutoramento inédita, Universidade de Lisboa / Universidade de São Paulo, 2018.

ANEXO: MAPAS



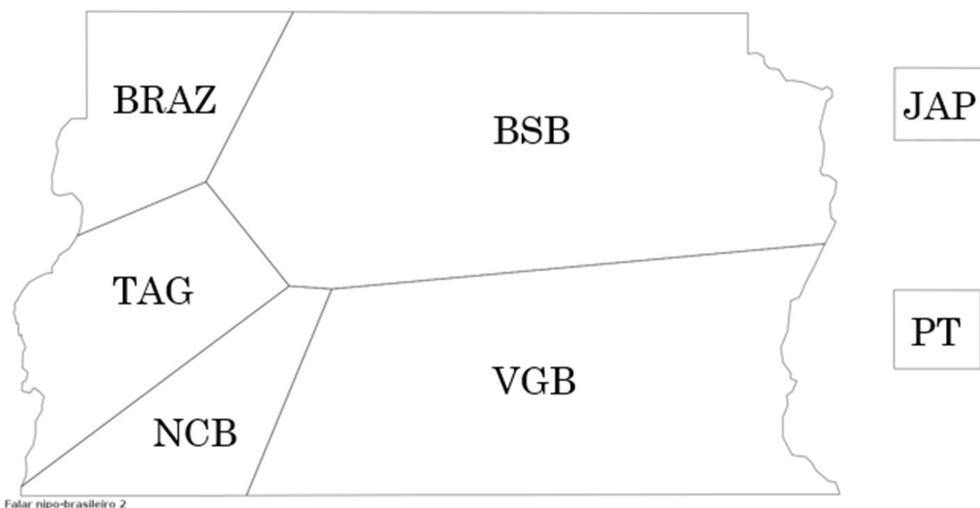
Mapa 1: Rede de pontos de inquérito de Takano (2013)

(Fonte: TAKANO, 2018.)



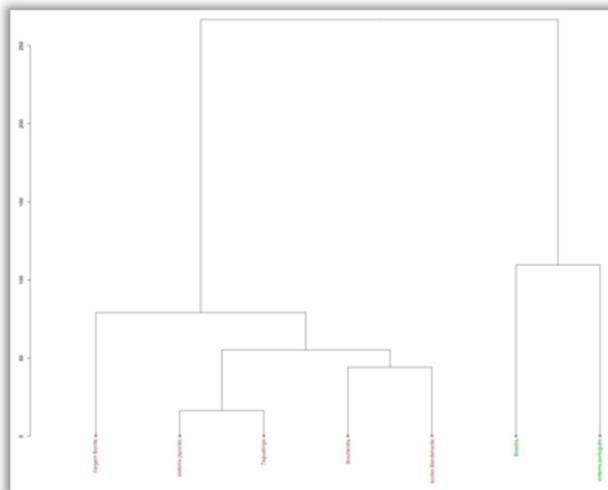
Mapa 2: Análise dendrográfica de Takano (2013)

Corpus: léxico, totalidade (97 conceitos); índice de similaridade: IRI; algoritmo: Ward; número de agrupamentos: 2. (Fonte: TAKANO, 2018.)



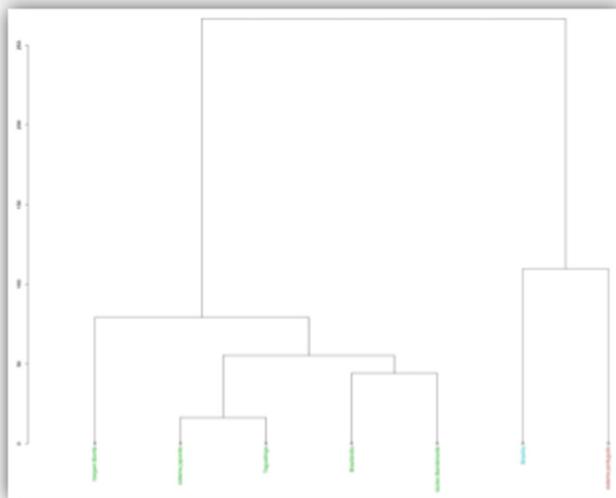
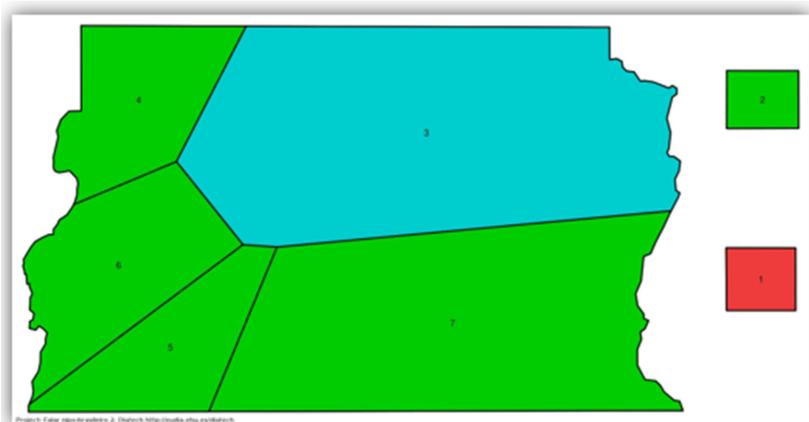
Mapa 3: Rede de pontos de inquérito de Takano (2013, 2018) modificada

Legenda: incluem-se os 5 pontos de inquérito de Takano (2013, 2018) e adicionam-se dois outros pontos de inquérito artificiais correspondentes ao sistema japonês e ao sistema português. Nomes dos pontos de inquérito: **BRAZ** = Brazlândia; **BSB** = Brasília; **NCB** = Núcleo Bandeirante; **TAG** = Taguatinga; **VGB** = Vargem Bonita. **JAP** = Japonês; **PT** = Português brasileiro.



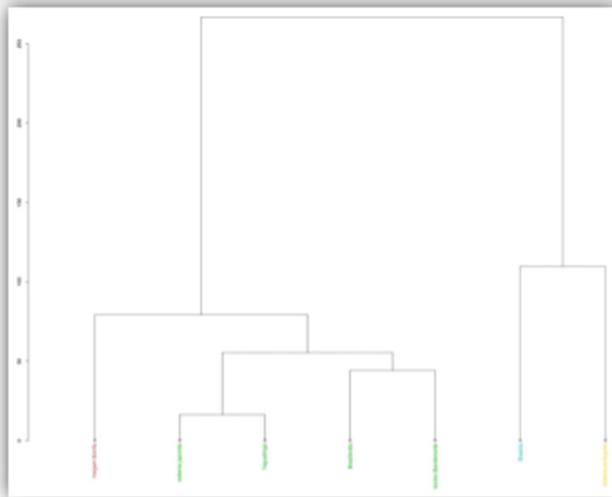
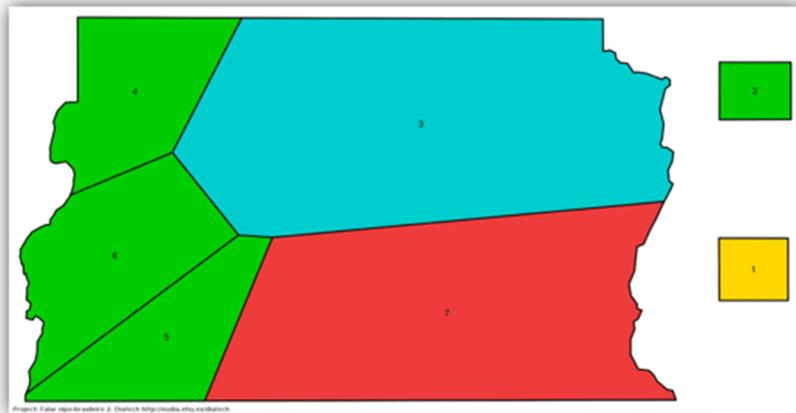
Mapa 4: Análise dendrográfica de Takano (2013) + JAP e PT

Corpus: léxico, totalidade (97 conceitos); índice de similaridade: IRI;
algoritmo: Ward; número de agrupamentos: 2.



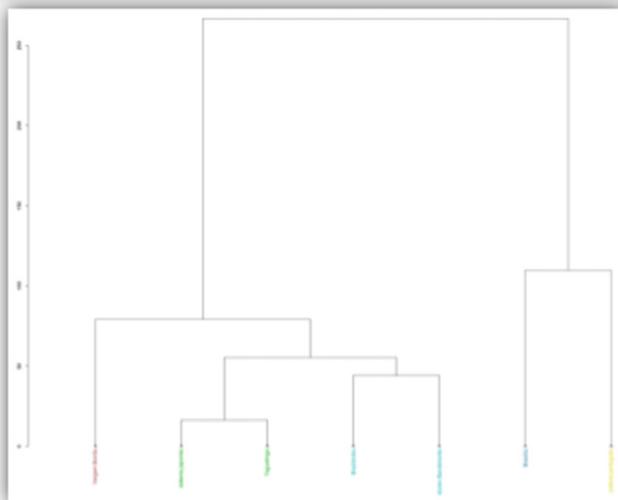
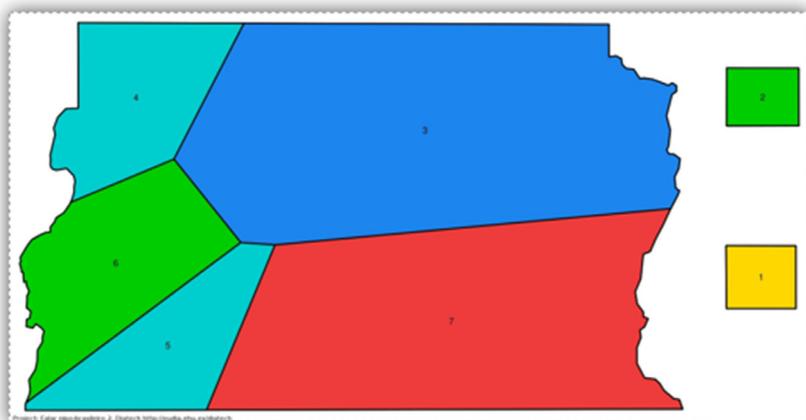
Mapa 5: Análise dendrográfica de Takano (2013) + JAP e PT

Corpus: léxico, totalidade (97 conceitos); índice de similaridade: IRI;
 algoritmo: Ward; número de agrupamentos: 3.



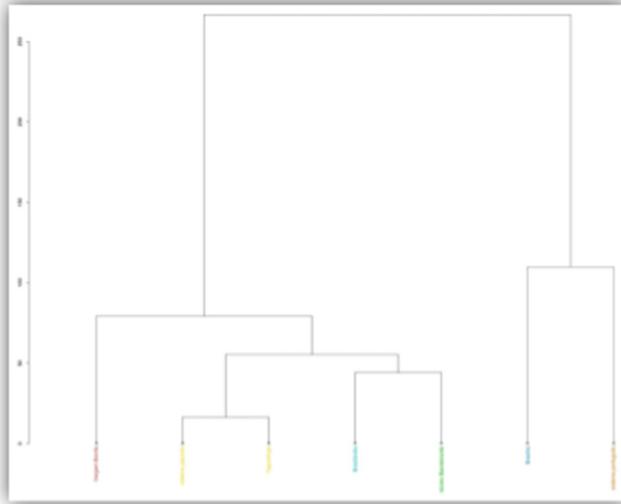
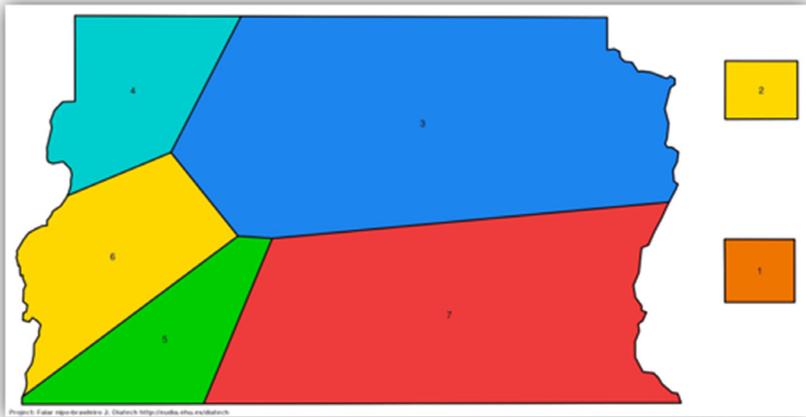
Mapa 6: Análise dendrográfica de Takano (2013) + JAP e PT

Corpus: léxico, totalidade (97 conceitos); índice de similaridade: IRI;
 algoritmo: Ward; número de agrupamentos: 4.



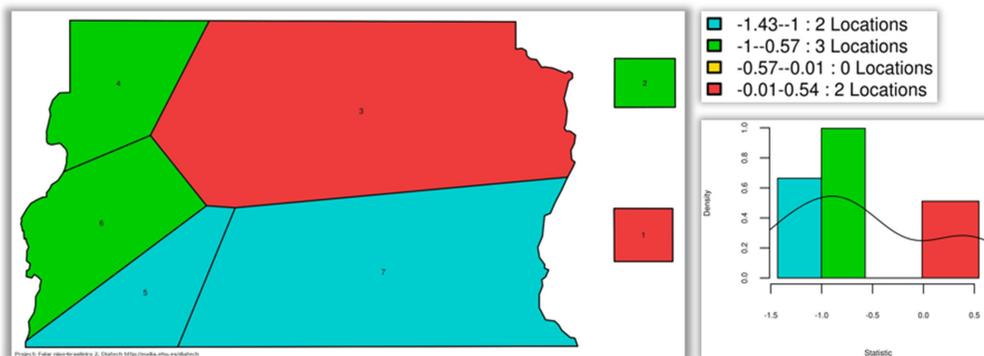
Mapa 7: Análise dendrográfica de Takano (2013) + JAP e PT

Corpus: léxico, totalidade (97/ conceitos); índice de similaridade: IRI;
algoritmo: Ward; número de agrupamentos: 5.



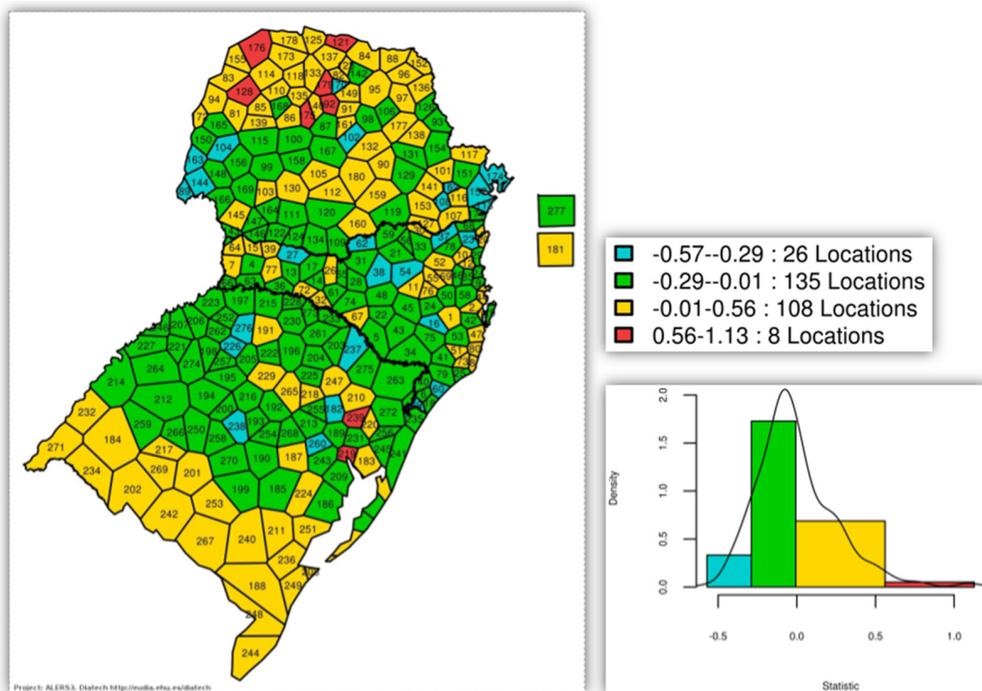
Mapa 8: Análise dendrográfica de Takano (2013) + JAP e PT

Corpus: léxico, totalidade (97 conceitos); índice de similaridade: IRI;
 algoritmo: Ward; número de agrupamentos: 6.



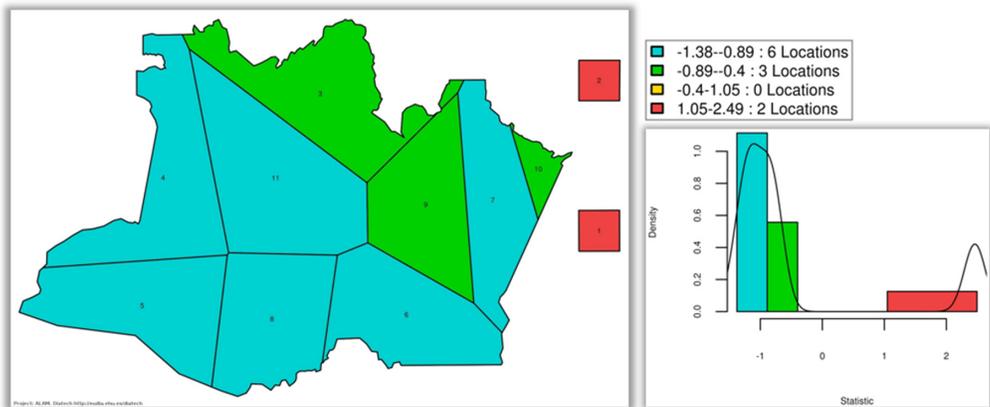
Mapa 9: Sinopse da *distribuição skewness* (coeficiente de assimetria de Fischer) de Takano (2013) + JAP e PT

Corpus: léxico, totalidade (97 conceitos); índice de similaridade: IRI; algoritmo de visualização: MINMWMAX 4-tuplo.



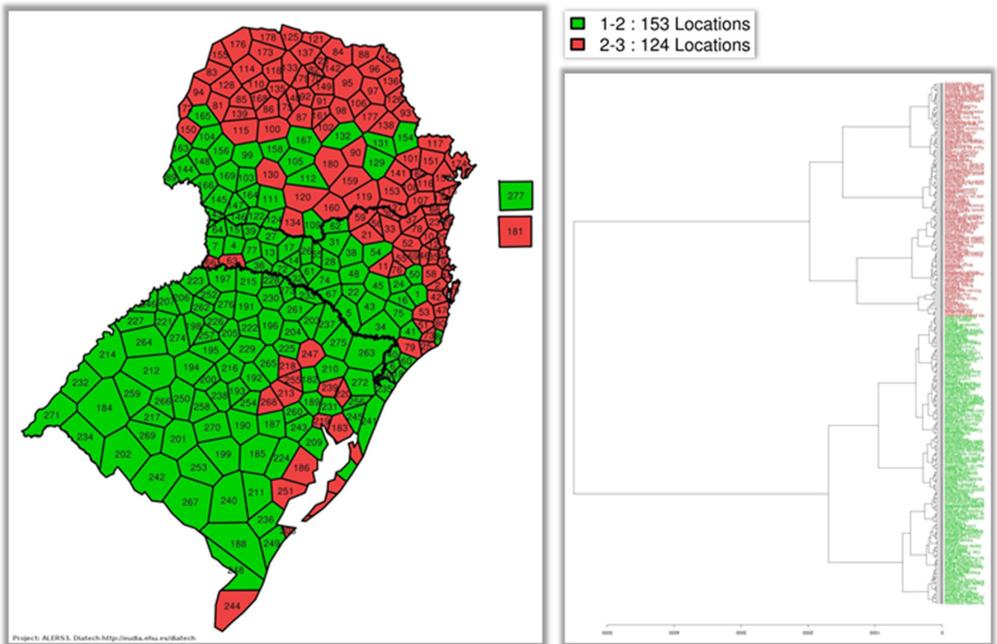
Mapa 10: Sinopse da *distribuição skewness* (coeficiente de assimetria de Fischer) do Atlas Linguístico-Etnográfico da Região Sul do Brasil + português padrão europeu e português padrão brasileiro

Corpus: léxico, totalidade (100 conceitos); índice de similaridade: IRI; algoritmo de visualização: MINMWMAX 4-tuplo. (Fonte: BRISSOS; SARAMAGO, 2019.)



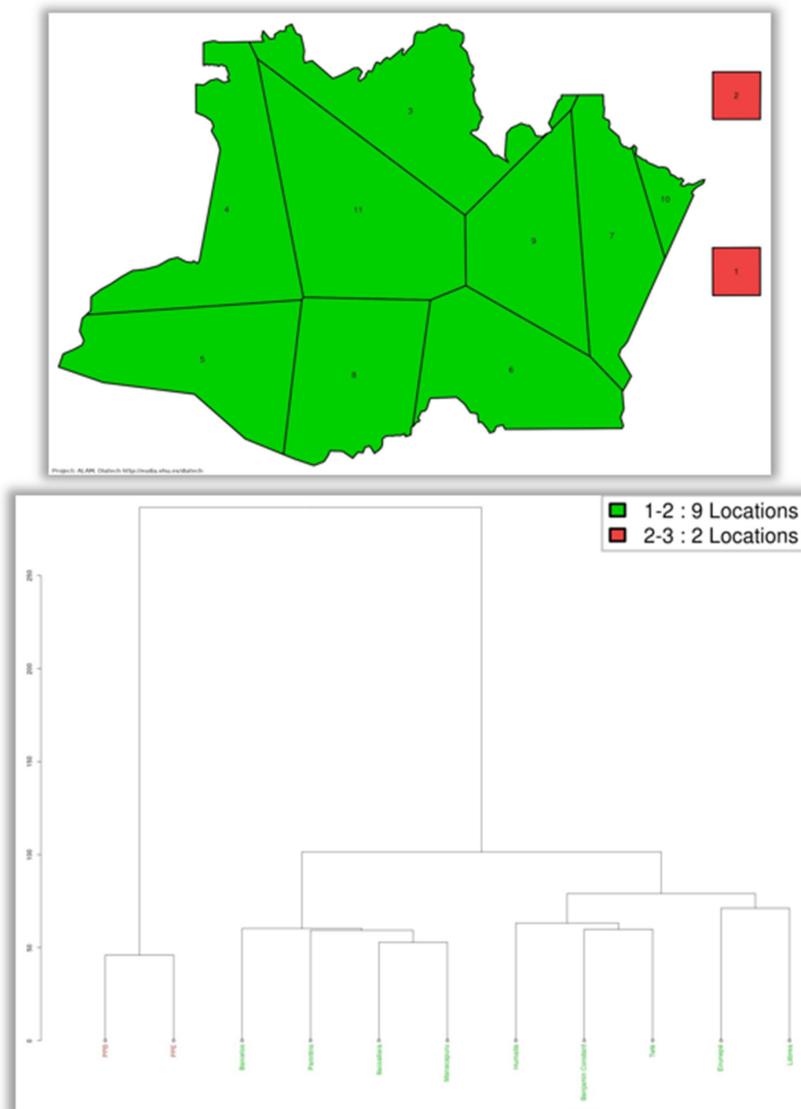
Mapa 11: Sinopse da *distribuição skewness* (coeficiente de assimetria de Fischer) do Atlas Linguístico do Amazonas + português padrão europeu e português padrão brasileiro

Corpus: léxico, totalidade (81 conceitos); índice de similaridade: IRI; algoritmo de visualização: MINMWMAX 4-tuplo. (Fonte: BRISSOS, 2021a.)



Mapa 12: Análise dendrográfica do Atlas Linguístico-Etnográfico da Região Sul do Brasil + português padrão europeu e português padrão brasileiro

Corpus: léxico, totalidade (100 conceitos); índice de similaridade: IRI; algoritmo: Ward; número de agrupamentos: 2. (Fonte: BRISSOS; SARAMAGO, 2019.)



Mapa 13: Análise dendrográfica do Atlas Linguístico do Amazonas + português padrão europeu e português padrão brasileiro

Corpus: léxico, totalidade (81 conceitos); índice de similaridade: IRI; algoritmo: Ward; número de agrupamentos: 2. (Fonte: BRISSOS, 2021a.)

*Recebido em 10 de junho de 2021.
Aprovado em 8 de setembro de 2021.*