

KUBATA

Prefácio

Olá! Bem-vindo à 1ª fase B da décima edição da Olimpíada Brasileira de Linguística!

Esperamos que esta edição, em casa, nos ajude a reconhecer e integrar as presenças de origem indígena, africana e sul-americana no Brasil e cultivar melhor nossa casa comum com todos os povos e os seres em geral.

Essa prova tem 12 questões de múltipla escolha, para serem resolvidas entre as 14:00 e as 20:00 do dia 18 de outubro de 2020. Você pode fazer a prova pelo aplicativo (celular, tablet, etc.) ou no browser do seu computador, durante as seis horas disponíveis.

Quando terminar a prova, você precisa clicar no botão "Finalizar", para enviar suas respostas aos nossos servidores. Esse envio precisa ser feito até as 19:59 do dia 18 de outubro.

Ao prosseguir, você concorda com o **compromisso de honra** de resolver os problemas **individualmente**, sem auxílio de colegas ou professores, e **sem consulta** à internet ou a qualquer outra fonte de pesquisa, confiando apenas na sua intuição, seu raciocínio e seu conhecimento de mundo para desvendar os fenômenos apresentados, cultivando uma cultura olímpica de valorização do conhecimento, da integridade e da autonomia.

Boa caminhada!

Esta prova contém problemas compostos por

Aleksejs Peguševs, Artur Corrêa, Cynthia Herkenhoff, Douglas Teixeira, Jade Yarden Steinmetz,
Janaina Weissheimer, João Henrique Fontes, Rafael Righetto, Roger Antunes e Vlad Neacșu

Além disso, ela foi editada, testada e revisada por

Artur Corrêa, Bruno L'Astorina, Cynthia Herkenhoff, Gustavo Palote, Eduardo Cardoso Martins, Felipe Petilo, João
Henrique Fontes, Pedro Martins Leão, Rafael Righetto, Rodrigo Pinto Tiradentes e Vlad Neacșu

#1 · Avião

Pilotos de avião do mundo todo, quando precisam soletrar, utilizam um sistema que assinala um nome para cada letra, o alfabeto fonético da ICAO (Organização da Aviação Civil Internacional). Esse sistema existe para eliminar possíveis equívocos na comunicação; quando alguém soletra uma palavra, mesmo que o som fique interrompido pela estática ou por alguma interferência na transmissão, ainda é possível compreendê-la, pois cada nome corresponde a uma única letra.

Cada avião possui uma sequência de letras para identificá-lo, como se fosse uma placa de carro, chamada *matrícula*. Ao se comunicar com a torre de controle, por exemplo, o piloto deve soletrar a matrícula do seu avião.



<https://youtu.be/UFv7YOVS6YE>

No vídeo acima, há uma compilação de trechos de cinco vôos. Abaixo está a lista dos órgãos de serviço de tráfego aéreo com os quais os pilotos se comunicam, com suas denominações na ordem em que aparecem no vídeo. Além disso estão também as matrículas dos aviões que aparecem no vídeo, em ordem aleatória:

*Controle Belo Horizonte, Torre Guarulhos, Controle São Paulo,
Torre Rio Branco, Rádio Rondonópolis.*

PR-MSO, PR-PLR, PT-VLX, PU-SJG, _____

Qual é a matrícula faltando?

- a) PT-OGU
- b) PR-OTZ
- c) PR-OTU
- d) PR-OTS
- e) PT-OGS

— Artur Corrêa, Douglas Teixeira



#2 · Băile Obeleului

Os estados-nação foram configurados com uma visão linguística em que cada estado corresponderia ao território dos falantes de uma língua. Raramente, no entanto, a realidade corresponde a este ideal. É o caso da Romênia, cuja região histórica da Transilvânia é habitada há séculos, simultaneamente, por falantes de romeno, de húngaro e de alemão. Hoje, essa situação tem algum grau de reconhecimento oficial, de forma que muitas cidades e vilas da região possuem dois nomes: um nome oficial em romeno e o nome correspondente em húngaro.

Abaixo, listamos dez desses lugares. Na primeira tabela, temos os nomes em romeno e, entre parênteses, o nome do distrito em que ela está contida. Em seguida, a lista dos nomes correspondentes em húngaro, em ordem aleatória.

Baia Mare (MM)	Chedia Mare (HR)
Băile Felix (BH)	Finteușu Mic (MM)
Băile Homorod (HR)	Ilva Mică (BN)
Băile Tușnad (HR)	Porumbenii Mari (HR)
Bocicoiu Mare (MM)	Sebeșu Mic (CJ)

Félixfürdő, Homoródfürdő, Kisfentős, Kisilva, Kissebes, Nagybánya,
Nagybocskó, Nagyalambfalva, Nagykedde, Tusnádfürdő

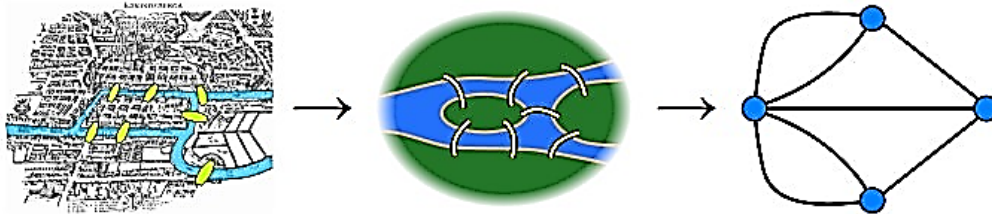
Quais das alternativas correspondem às seguintes cidades de nome romeno: **Băile Herculane** (CS), **Chendu Mic** (MS), e **Mireșu Mare** (MM)?

- a) Nagyherkulane, Kiscsend, Nyíresfürdő
- b) Nagyherkules, Kiskend, Nyíresfürdő
- c) Nagyherkules, Kiskend, Nagynyíres
- d) Herculanefürdő, Kiscsend, Nagynyíres
- e) Herkulesfürdő, Kiskend, Nagynyíres

— Vlad Neacșu

#3 · As Sete Pontes de Königsberg

O matemático suíço Leonhard Euler ficou conhecido em 1736 por resolver o problema das Sete Pontes de Königsberg, representando locais da cidade como vértices e pontes como arestas, o que deu origem à Teoria de Grafos.



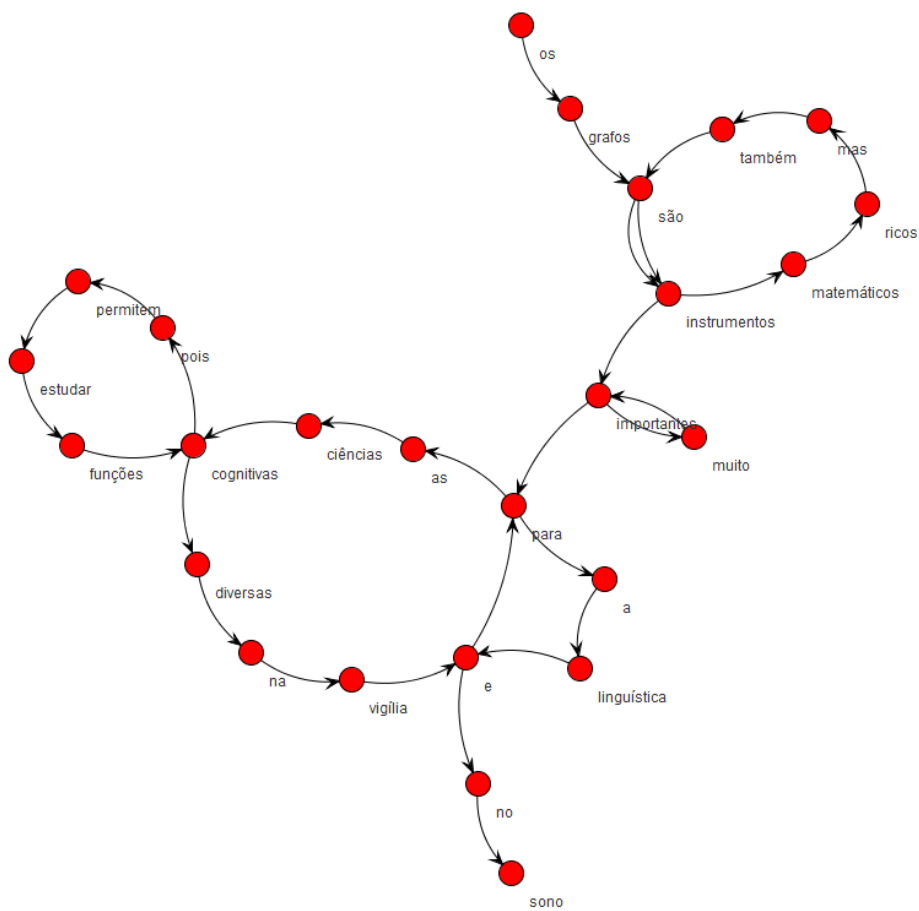
Atualmente, linguistas aplicam essa teoria para investigar a relação entre a trajetória de palavras em um discurso. Em geral, discursos que geram mais conexões aparecem em indivíduos com pontuações mais altas em avaliações cognitivas. Ou seja, os grafos linguísticos podem prever certas performances cognitivas.

Na análise do grafo, podemos utilizar diferentes parâmetros, com interpretações linguísticas distintas. Por exemplo, para medir a **conectividade** do discurso (um índice da organização do pensamento), podemos usar parâmetros como o *grau total médio* (ATD) e o *maior componente fortemente conexo* (LSC). Para avaliar a **recorrência** do discurso (isto é, a repetição de palavras ou estruturas), pode-se usar o *número de arestas repetidas* (RE) e de *arestas paralelas* (PE).

O grau de um vértice é o número de arestas (tanto entrando quanto saindo) ligadas a ele. O grau total médio (ATD) é a média entre os graus dos diferentes vértices. No caso de um grafo direcionado, como os que aparecem nessa análise linguística, um componente conexo equivale a um ciclo fechado. Assim, o maior componente fortemente conexo (LSC) é o maior *ciclo fechado* do grafo, isto é, o maior subgrafo em que todas as palavras estejam mutuamente conectadas por uma trajetória direcionada: existe caminho tanto de a para c quanto de c para a, por exemplo, $a \rightarrow b \rightarrow c \rightarrow a$. O número de arestas repetidas (RE) é o número de ocasiões em que um par de vértices é ligado por mais de uma aresta com a mesma direção ($a \rightarrow b$); enquanto o número de arestas paralelas (PE) conta pares de vértices ligados por arestas em direções opostas ($a \leftrightarrow b$).

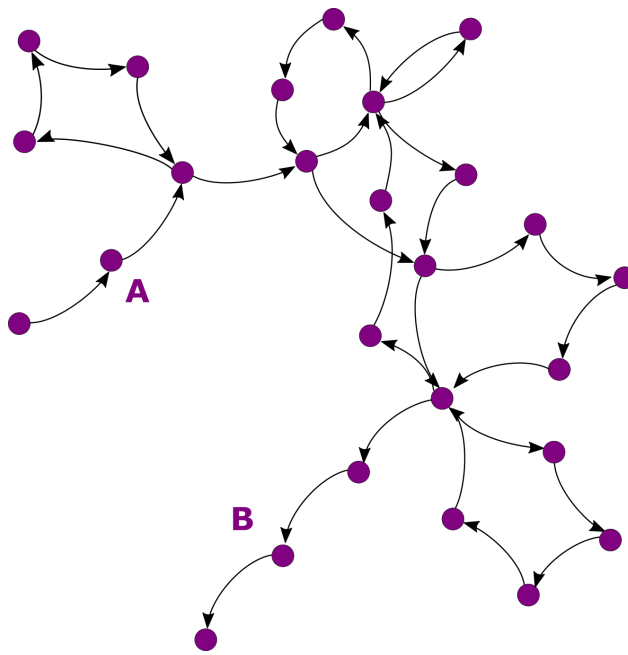
Como exemplo, veja o grafo abaixo junto com a sentença que o gerou. Este grafo tem 26 vértices e 31 arestas e tem os seguintes parâmetros: ATD = 2,38, LSC = 14, PE = 1 e RE = 1.



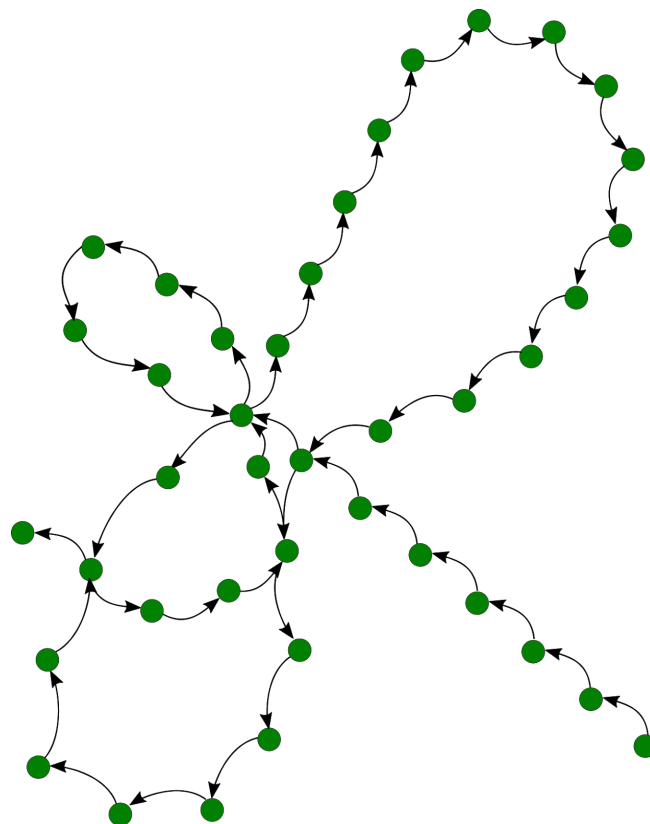


Os grafos são instrumentos matemáticos ricos, mas também são instrumentos importantes, muito importantes, para a linguística e para as ciências cognitivas, pois permitem estudar funções cognitivas diversas na vigília e no sono.

A seguir temos dois grafos retirados de histórias criadas por crianças, usados em uma pesquisa. Neste caso, ocultamos as palavras dos grafos, mas mantemos o texto correspondente abaixo de cada um.



Era um dia ensolarado. E também, dia de aniversário. O aniversário do Bob! Para comemorar esse aniversário, os pais de Bob convidaram alguns amigos para ir a casa deles, para fazer uma festa!



Essa história é sobre um menino e uma avó que ficou cuidando do garoto enquanto os pais estavam fora então eles saíram e foram em vários lugares como a quadra de tênis depois foram dar uma longa caminhada pelo bosque com uma trilha de areia

Sobre os dois grafos acima, marque a alternativa **incorreta**:

- a) No grafo 1, existem apenas dois vértices de grau 6, correspondendo às palavras “aniversário” e “para”.
- b) No grafo 1, o vértice A corresponde a “um” e o vértice B corresponde a “uma”.
- c) No grafo 1, os vértices "o" e "aniversário" são ligados por arestas paralelas (PE) e, portanto, formam um componente conectado.
- d) No grafo 2, o parâmetro LSC vale 33.
- e) No grafo 2, o primeiro vértice que possui grau maior que 2 corresponde à palavra “uma”.

— *Janaina Weissheimer, Cynthia Herkenhoff*

#4 · Numerus Cisterciensis

Abaixo são dados alguns numerais escritos em uma notação especial, usada principalmente pelos monges cistercienses entre os séculos XIII e XV. Os mesmos números estão escritos com algarismos indo-arábicos, mas em ordem aleatória.



1486	1801	1912	2421	9999
------	------	------	------	------

Que número é representado pelo símbolo abaixo?



- a) 1789
- b) 7891
- c) 8971
- d) 9178
- e) 9817

— Vlad Neacșu



#5 · Bopomofo

O Bopomofo, também chamado de *Zhuyin Fuhao* (注音符號, lit. *símbolos fonéticos*), é um sistema de escrita fonética para o Mandarim, utilizado principalmente em Taiwan. No áudio abaixo são faladas seis palavras em Mandarim com suas transcrições em ordem aleatória em Bopomofo:

<https://drive.google.com/file/d/1VYbWYjZKCCkLohMt-d6I34jzAsG7AwZ5/view?usp=sharing>

ㄌㄨˋ, ㄎㄨㄛˋ, ㄌㄛㄝˊ, ㄉㄨㄛˋ, ㄎㄨㄟˊ, ㄌㄨㄟˊ

Quais são as três primeiras palavras faladas no áudio?

- a) ㄌㄛㄝˊ, ㄌㄨㄟˊ, ㄎㄨㄛˋ
- b) ㄌㄨˋ, ㄉㄨㄛˋ, ㄎㄨㄟˊ
- c) ㄌㄨˋ, ㄌㄨㄟˊ, ㄎㄨㄟˊ
- d) ㄌㄛㄝˊ, ㄉㄨㄛˋ, ㄎㄨㄛˋ
- e) ㄌㄛㄝˊ, ㄌㄨㄟˊ, ㄎㄨㄟˊ

— João Henrique Fontes

#6 · Hauçá

A língua **hauçá**, pertencente ao ramo chádico da família afroasiática, é uma das mais importantes da África, sendo falada como língua nativa pelo povo Hauçá e como *lingua franca* por aproximadamente 100 milhões de pessoas na África Ocidental, principalmente na Nigéria. Há uma grande produção literária nessa língua, notadamente sua poesia, na qual o ritmo dos versos é bem importante.

Abaixo estão cinco versos de poesia hauçá escritas no alfabeto Boko (de base latina), sendo 4 seguindo um padrão rítmico e 1 de outro padrão. Identifique esse outro.

Nota: Os tons das vogais hauçá foram omitidos nesta questão, o que não interfere na sua resolução. O w sobrescrito representa uma labialização da consoante anterior, ou seja, k^w é apenas uma consoante. Vogais duplicadas representam vogais longas.

- a) Allaahu yaa yi umuñnii duk ga alummaa
- b) Dan jaahilii dabbaa fa nee tun bay mutu
- c) Wallaahi in har mun tsayaa mun bincikaa
- d) Zaa mui tunaanii mai yawaa gun zuuciyaa
- e) Doomin a lootan nan da jaahilcii k^wañai

— João Henrique Fontes



#7 · Tamahaq

As línguas dos povos chamados genericamente de *tuaregues*, que habitam toda a extensão do Saara, são um exemplo da dificuldade dos linguistas de se estabelecer uma divisão clara do que é uma língua e do que é um dialeto. O **tamahaq**, ou também conhecido como tahagart, é um dos muitos dialetos falados pelo povo Imuhar, no sul da Argélia e da Líbia e no norte do Níger. Seguem abaixo algumas palavras em tamahaq e suas traduções para o português:

tamart	<i>barba</i>	timaren	<i>barbas</i>
turift	<i>barco</i>	turifen	<i>barcos</i>
amawey	<i>chefe</i>	imaweyen	<i>chefes</i>
abaraḍ	<i>menino</i>	ibaraḍen	<i>meninos</i>
tisut	<i>tosse</i>	tisuten	<i>tosses</i>
ige	<i>ação</i>	igetan	<i>ações</i>
aghaba	<i>rédea</i>	ighabatan	<i>rédeas</i>
	<i>gafanhoto</i>	tihwalen	<i>gafanhotos</i>
tabaraṭ	<i>menina</i>		<i>meninas</i>
ini	<i>cor</i>		<i>cores</i>
asulmey	<i>peixe</i>	isulmeyen	<i>peixes</i>
	<i>acordo</i>	iharaḍen	<i>acordos</i>
akasa	<i>fornagem</i>	ikasatan	<i>fornagens</i>
abaraqqa	<i>caminho</i>	ibaraqqatan	<i>caminhos</i>
ahar	<i>leão</i>	iharen	<i>leões</i>
taslaft	<i>corvo</i>	tislafen	<i>corvos</i>

Nota: ḍ e ṭ são as chamadas versões enfáticas de ‘t’ e ‘d’, pronunciadas com uma espécie de “impulso” da garganta; sh é como ‘x’ em ‘xícara’; gh é como ‘rr’ em ‘carro’, com as cordas vocais vibrando; w e y são consoantes com o mesmo som do inglês ‘water’ e ‘yes’.

Quais são as palavras que, de cima para baixo, preenchem as lacunas?

a) tihwalt, tabaraṭan, initan, iharaḍ

b) tahwal, tabaraṭen, initan, aharaḍ



c) tahwalt, tibaratan, iniren, iharađ

d) tahwalt, tibaraten, initan, aharađ

e) tahwalt, tibaraten, inien, aharađ

— *Rafael Righetto*



#8 · Hebraico Cursivo

Abaixo estão algumas palavras em hebraico cursivo ashkenazi e sefaradi respectivamente, com transliteração em alfabeto latino e tradução para o português:

cursivo ashkenazi	cursivo sefaradi	transliteração	
מנגן	מנגן	menagen	(ele, ela) toca
להעריך	לעריך	lehaarich	valorizar
שותפה	שותפה	shutafah	sócia
זמן	זמן	zman	tempo
עכשיו	עכשיו	achshav	agora
פיצויים	פיצויים	pitsuyim	compensação

Nota: *ch* tem um som parecido com 'rr' em 'carro'; *sh* é como 'x' em 'xícara'.

Como se escreve **gashum** (chuvoso), em hebraico cursivo ashkenazi, e **tsipor** (pássaro), em hebraico cursivo sefaradi?

a) זיפור, פהון

b) גושים, גשום

c) זיפור, גשום

d) זיפור, גשון

e) גאום, ציבור

— Jade Yarden Steinmetz, Rafael Righetto



#9 · Mboi-tata

Mboi-tata, ou *Boitató*, é uma lenda de origem indígena do folclore brasileiro. Uma de suas versões conta que, durante um período de noites sem fim e muita chuva, a **sucuri**, a maior cobra do mundo, acordou e saiu para comer os animais, mas só comia os olhos e nada mais. Por ter comido tantos olhos, a sucuri foi se tornando transparente, transparente, clareada pelas milhares de luzes dos olhos que comeu. E foi se tornando tão clara que virou chama, um fogo amarelo azulado. A partir de então, ficou conhecida por um novo nome: **mboi-tata**, que significa *cobra-de-fogo*.

O guarani mbya é um dos dialetos do guarani, língua da família tupi. No Brasil, estão distribuídos principalmente na Região Sul. Na tabela estão palavras em guarani mbya e, abaixo, as traduções em português, em ordem aleatória.

jaxy	mboi guaxu	tata rataxĩ
jaxy-tata	py'a	ye guaxu
jaxy endy	py'a guaxu	yvy rataxĩ
kuã guaxu	tata	yvy porã
kuã regua	tata endy	yy

água, anel, corajoso, dedão, estrela, fígado/coração, fogo, fumaça, grávida, lua, luar, luz do fogo, poeira, sucuri, terra boa

Nota: x é como 'tch' em 'tchau'; j é como 'dj' em 'djangô'; y é uma vogal intermediária entre 'i' e 'u'; ' é a pausa glotal, o '-' em 'ã-ã'. Todas as palavras da tabela são oxítonas.

Faça as correspondências entre as palavras. Depois disso, responda: como se diz, em guarani mbya, **tranquilo** e **nevoeiro**?

- a) py'a porã, yy rataxĩ
- b) jaxy rataxĩ, yvy guaxu
- c) kuã endy, yy tata
- d) tata py'a, yvy endy
- e) yy guaxu, jaxy kuã

— Artur Corrêa



#10 · CJK

O Mandarim moderno, assim como outros dialetos sinóticos como o Wu (Shanghainês) e o Yue (Cantonês), é derivado do **Chinês Médio** (CM), falado nos tempos das dinastias Tang e Song. Devido à longa centralidade cultural, tecnológica e econômica da China, outras línguas não derivadas do CM, como o Japonês, o Coreano e o Vietnamita, receberam diversos empréstimos dessa língua. Com o passar do tempo, os empréstimos continuaram a evoluir de forma independente, de forma que são raros os casos em que há correspondências regulares de sons. Ainda assim, comparando as línguas, é possível descobrir algumas características do CM.

Abaixo, temos algumas palavras em mandarim, coreano e japonês, escritas em transcrições latinas que representam seus sons.

mandarim	coreano	japonês	
nao.po	no.po	nō.ha	<i>onda cerebral</i>
wun.hua	mun.hwa	bun.ka	<i>cultura</i>
wei.xien	ui.hǒm	ki.ken	<i>perigo</i>
nan.xing	nam.sǒng	nan.sē	<i>masculino</i>
pa.nien	pal.nyǒn	hati.nen	<i>oito anos</i>
xien.sheng	sǒn.sěng	sen.sē	<i>professor, senhor</i>
wu.hao	o.ho	go.gō	<i>quinto</i>
xin.wun	sin.mun	sin.bun	<i>notícia, jornal</i>
fa.ming	pal.myǒng	hatsu.mē	<i>invenção</i>

A partir dessa tabela, podemos formular as seguintes hipóteses:

1. O alongamento das vogais no japonês é resultado de apenas um tipo de terminação de sílaba do CM.
2. O **x** do mandarim tem pelo menos duas origens distintas no CM.
3. O **w** do mandarim tem apenas uma origem no CM.
4. O **h** do japonês é uma modificação de apenas um som do CM.

Nota: o macron (-) sobre a vogal indica que a vogal é longa; ǒ tem um som próximo de ‘ó’ em ‘pó’; ǔ é como ‘é’ em ‘café’; h é como ‘r’ em ‘rato’; ng representa uma única consoante.



Estão provavelmente corretas:

- a) Apenas 2 e 3
- b) Apenas 2 e 4
- c) Apenas 1 e 4
- d) Apenas 1 e 2
- e) Apenas 1, 3 e 4

— *João Henrique Fontes*

#11 · Wa'â'ke

A língua tukano, língua-tema da edição Ye'pâ-masa da OBL, é falada por cerca de 5 a 6 mil pessoas no noroeste da Amazônia, especialmente no município de São Gabriel da Cachoeira, onde é uma das línguas oficiais, e em parte da Colômbia.

Observe, a seguir, algumas sentenças Tukano, com suas traduções para o português:

mi'i akawerérã yi'î pakó yaá kométi yahawã

Eu vi que seus parentes roubaram a panela da minha mãe

masí diâyirã yi'î pakí toâasã

Eu percebi que os cachorros do homem acabaram de morder meu pai

yi'î pîrôrã mi'i bi'î ba'âasã

Eu percebi que as minhas cobras acabaram de comer seu rato

mi'i pakó yi'î yaá yukîsipi yahasami

Eu percebi que sua mãe rouba minha canoa

mi'i bi'î mi'i yaá sêragá ba'âmi

Eu vi que seu rato come seu abacaxi

masí pîrôrã mi'i akaweré toâsama

Eu percebi que as cobras do homem mordem seu parente

mi'i bi'îrã yi'î diâyi toâama

Eu vi que os seus ratos acabaram de morder meu cachorro

mi'i pakí yi'î yaá kométi-yahaami

Eu vi que o seu pai acabou de roubar minha panela

yi'î akaweré yi'î yaá sêragá ba'âsami

Eu percebi que meu parente come meu abacaxi

mi'i diâyirã masí yaá yukîsipi toâwã

Eu vi que seus cachorros morderam a canoa do homem

Nota: ´, ʌ, ~ representam tom ascendente, tom alto e som nasal, respectivamente. ' é a pausa glotal, o '- em 'ã-ã'. i é uma vogal intermediária entre 'i' e 'u'.

Como se diz em tukano, respectivamente:

Eu vi que o homem cortou sua lenha

Eu percebi que o seu parente acabou de comer meus peixes.

a) masí mi'i pekâ direáwĩ; mi'i akaweré yi'î wa'îrã ba'âasã



- b) masí mi'i yaá pekâ direáwã; mi'i akaweré yi'î wa'îrã ba'âisã
- c) masí mi'i pekâ direáwĩ; mi'i akaweré yi'î wa'îrã ba'âasĩ
- d) masí mi'i yaá pekâ direáwĩ; mi'i akaweré yi'î wa'îrã ba'âasĩ
- e) masí mi'i pekâ direáwã; mi'i akaweré yi'î wa'îrã ba'âisã

— Roger Antunes, João Henrique Fontes

#12 · Citos Gadijumos

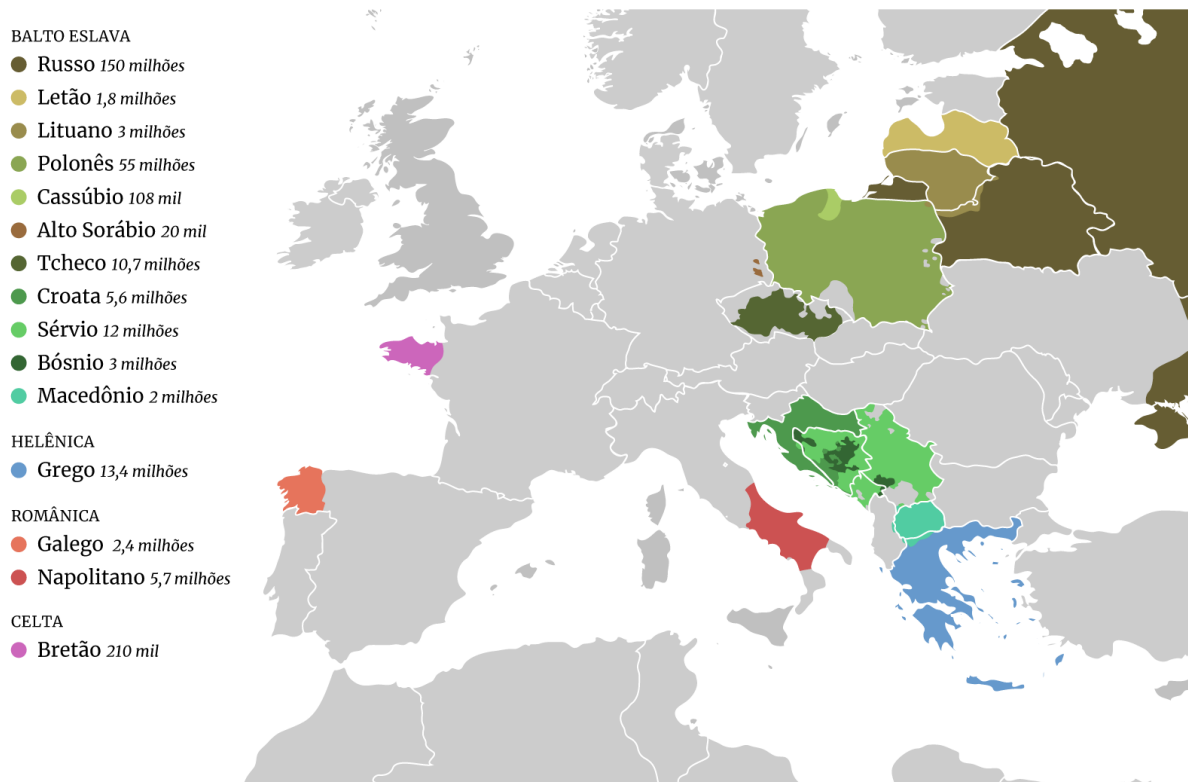
Seguem os nomes dos primeiros quatro meses do ano em várias línguas europeias:

Bósnio	januar	februar	mart	april
Bretão	miz Genver	miz C'hwevrer	miz Meurzh	miz Ebrel
Croata	siječanj	veljača	ožujak	travanj
Tcheco	leden	únor	březen	duben
Grego	Ianouários	Febrouários	Mártios	Aprílios
Letão	janvāris	februāris	marts	aprīlis
Lituano	sausis	vasaris	kovas	balandis
Napolitano	jennaro	frevaro	màrzo	abbrile
Russo	janvar'	fevral'	mart	aprel'
Sérvio	januar	februar	mart	april
Alto Sorábio	wulki róžk	mały róžk	nalětnik	jutrownik

Todas essas línguas são da família indo-europeia, como o português, mas pertencem a subfamílias diferentes dentro dela. Algumas delas possuem o status de língua nacional, ou seja, são língua oficial de algum país independente. Outras são línguas minoritárias, de expressão regional, com pouco reconhecimento político – como o caso do Cassúbio, falado principalmente na região polonesa da Pomerânia, e o Alto Sorábio, falado principalmente na região alemã da Lusácia.

Para contextualizar, veja um mapa com a distribuição geográfica, a família e o número de falantes dessas línguas:





A seguir, uma tabela com a porcentagem de cristãos católicos, ortodoxos e protestantes nos territórios onde as línguas acima são faladas:

Território	Cat (%)	Ort (%)	Prot (%)
Bósnia	15	30	0
Bretanha	50	0	0
Croácia	86	4	0
Galícia	82	1	1
Grécia	0	90	3
Itália	67	3	1
Letônia	25	19	35
Lituânia	77	5	0
Lusácia	58	0	36
Macedônia	1	64	0
Pomerânia	92	0	0
Polônia	87	1	0
Rússia	0	43	0
Sérvia	5	85	1
Tchéquia	10	1	1

Quais são os nomes para:

- Janeiro, em macedônio
- Fevereiro, em galego
- Março, em cassúbio
- Abril, em polonês?

a) januari, febreiro, marc, apriel

b) sečnik, choiva, marc, apriel

c) sečnik, febreiro, strëmiannik, kwiecień

d) sečnik, choiva, strëmiannik, apriel

e) januari, febreiro, strëmiannik, kwiecień



