

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

Artur Corrêa Souza, 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 Souza, 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/UFv7YQVS6YE>

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 Souza, Douglas Teixeira

Resposta: B

O mais provável é que o piloto identifique seu avião, falando a matrícula, logo no início da comunicação, assim como normalmente apresentamos nosso nome logo no início da comunicação com alguém desconhecido. Podemos perceber que após o piloto mencionar o órgão de serviço de tráfego aéreo, ele fala uma sequência de nomes. Na maioria dos vídeos, o órgão repete essa sequência de nomes. Essa sequência de nomes deve ser a matrícula do avião.



Pelas quatro matrículas que conhecemos, é possível ver que (i) todas começam com o mesmo nome – o nome associado à letra P, aparecendo ao todo seis vezes; (ii) os nomes das letras L e S aparecem duas vezes, e o nome da letra R aparece três vezes; (iii) a matrícula PR-PLR é bem importante pois possui duas letras repetidas além de L que aparece em outra matrícula:

PR-MSO

PR-PLR

PT-VLX

PU-SJG

A parte mais difícil do problema é a acuidade auditiva, perceber os nomes que os pilotos falam. Fica mais fácil depois das considerações acima.

Na ordem em que aparecem, podemos ouvir:

- 1) Controle Belo Horizonte, **Papa Uniforme Sierra Juliete Golfe**
Papa Uniforme Sierra Juliete Golfe [...]
- 2) Torre Guarulhos, **Papa Romeu Maike Sierra Oscar** [...]
- 3) Boa noite, **Papa Tango Vitor Lima Éks-rei**
Papa Tango Vitor Lima Éks-rei, Controle São Paulo [...]
- 4) Torre Rio Branco, bom dia, **Papa Romeu Oscar Tango Zulu**
Papa Romeu Oscar Tango Zulu, torre [...]
- 5) Coordenação em Rondonópolis, **Papa Romeu Papa Lima Romeu** [...]

Os nomes do alfabeto da ICAO não são, na verdade, os transcritos acima. Os nomes foram pensados para falantes de inglês e são portanto escritos de outra forma. Maike se escreve Mike; Juliete se escreve Juliett; Golfe é Golf; e o misterioso Éks-ray é X-ray (raio-x). Perceber isso não era necessário para resolver o problema. Os falantes de português, assim como de qualquer outra língua, fazem variações nas palavras para acomodá-las de forma mais confortável na sua própria língua, gerando Juliete em vez de Juliet, por exemplo. Vamos manter aqui a escrita aproximada em português.

Podemos perceber, então, que PR-PLR é a última matrícula pelas duas palavras repetidas. Assim P = Papa, R = Romeu e L = Lima. Vemos que os nomes associados às letras **começam com a letra associada** (com exceção de Éks-rei, explicado acima). Fica fácil então associar as outras matrículas.

- | | |
|---------------------------|------------------------------------|
| 1) Belo Horizonte: PU-SJG | Papa Uniforme Sierra Juliete Golfe |
| 2) Guarulhos: PR-MSO | Papa Romeu Maike Sierra Oscar |
| 3) São Paulo: PT-VLX | Papa Tango Vitor Lima Éks-rei |
| 4) Rio Branco: _____ | Papa Romeu Oscar Tango Zulu |
| 5) Rondonópolis: PR-PLR | Papa Romeu Papa Lima Romeu |

A matrícula que falta é a de Rio Branco. As letras das palavras Papa, Romeu, Oscar e Tango podem ser descobertas pela inicial ou pela repetição em outras matrículas: P R O T. Sobram, então, as alternativas

- b) PR-OTZ
- c) PR-OTU
- d) PR-OTS

Para não restar dúvidas, podemos eliminar com certeza a alternativa (c) pois U é falado na 1ª matrícula (Uniforme); e eliminar também a alternativa (d) pois S é falado na 1ª e 2ª matrícula (Sierra). A matrícula correta só pode ser portanto a **B**.



#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)
Bocicioiu 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



Resposta: E

A partir dos nomes próprios, conseguimos fazer rapidamente algumas associações:

Băile Felix	Félix-fürdő
Băile Homorod	Homoród-fürdő
Băile Tuşnad	Tusnád-fürdő

Isso permite entender que **băile** corresponde a **-fürdő**. De fato, ambos significam “banho” ou “spa”, referindo-se a cidades de águas – equivalente a nomes brasileiros como *Águas de São Pedro* ou *Águas de Lindóia*.

Uma outra cidade que poderia ser encontrada dessa forma é:

Finteuşu Mic	Kis-fentős
--------------	------------

o que permitiria relacionar as outras cidades húngaras com o mesmo prefixo.

Sebeşu Mic	Kis-sebes
Ilva Mică	Kis-ilva

O que permite entender o segundo elemento: **mic/mică** correspondente a **kis-**. Ambos significam “pequeno/pequena”.

Com isso, sobram quatro nomes:

Baia Mare	Nagy-bánya
Chedia Mare	Nagy-kede
Bocicoiu Mare	Nagy-bocskó
Porumbenii Mari	Nagy-galambfalva

O terceiro elemento: **mare/mari** corresponde a **nagy-**. Ambos significam “grande/grandes”. Além disso, essa última sequência indica algumas acomodações fonológicas entre as línguas, como o **ch-** romeno correspondendo a **k-** em húngaro.

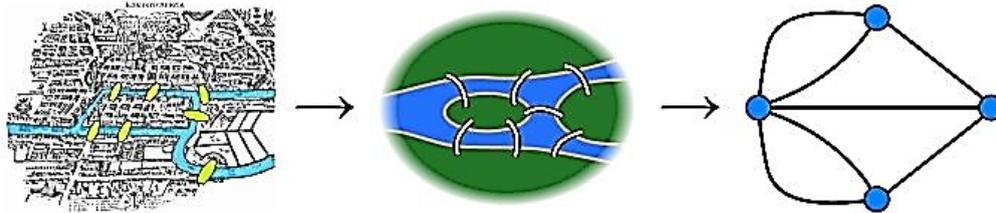
Analisando as alternativas, sabemos que temos que ter **-fürdő** na primeira palavra, **kis-** na segunda e **nagy-** na terceira. Isso nos deixa entre D e E. Pela transformação **ch-k** mencionada anteriormente, vemos que ‘chendia’ deve se tornar ‘kend’ e não ‘csend’. Uma outra forma de analisar seria entender que ‘herculane’ é uma forma adjetival própria do romeno, referindo-se ao nome ‘hercules’; a língua húngara não usaria essa derivação romena, usando o nome do herói na sua grafia, ‘herkules’. Em ambos os casos, concluímos que a alternativa correta é E.

Por fim, o título do problema mostra uma cidade fictícia da OBL (lida em romeno como “obele”); a forma **-ului** é uma forma de marcar posse, explorada no problema 2 da fase 2 da edição Yora. Assim, o título se traduz literalmente como “banhos da OBL”.



#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.

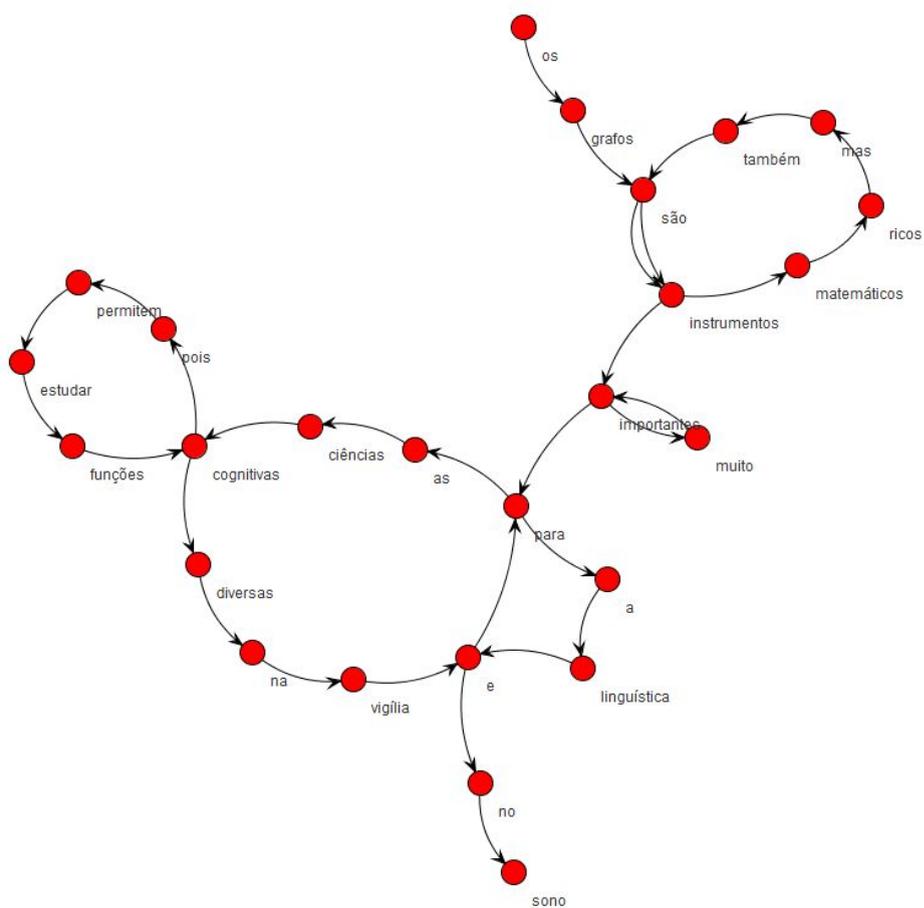


Atualmente, linguistas aplicam essa teoria para investigar a relação entre a trajetória de palavras em um discurso. Em geral, discursos que geram mais conexões aparecem em indivíduos com pontuações mais altas em avaliações cognitivas. Ou seja, os grafos linguísticos podem prever certas performances cognitivas.

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).

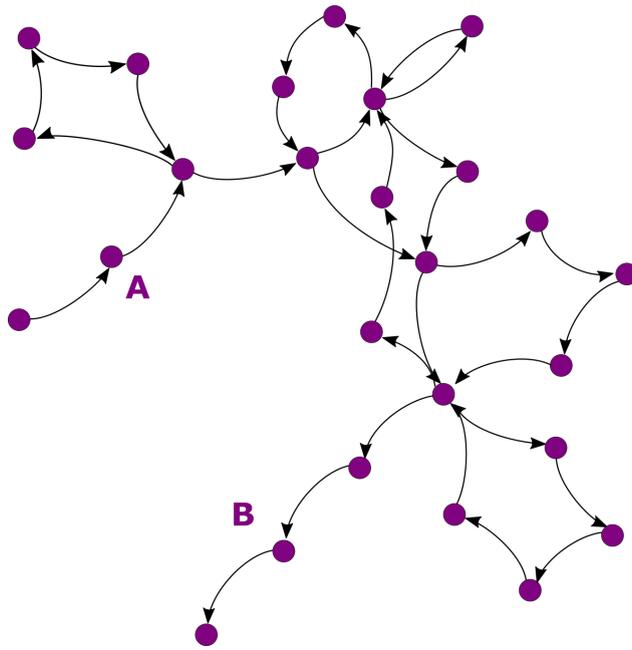
O grau de um vértice é o número de arestas (tanto entrando quanto saindo) ligadas a ele. O grau total médio (ATD) é a média entre os graus dos diferentes vértices. No caso de um grafo direcionado, como os que aparecem nessa análise linguística, um componente conexo equivale a um ciclo fechado. Assim, 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 ($a \rightarrow b$ e $b \rightarrow a$).

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.

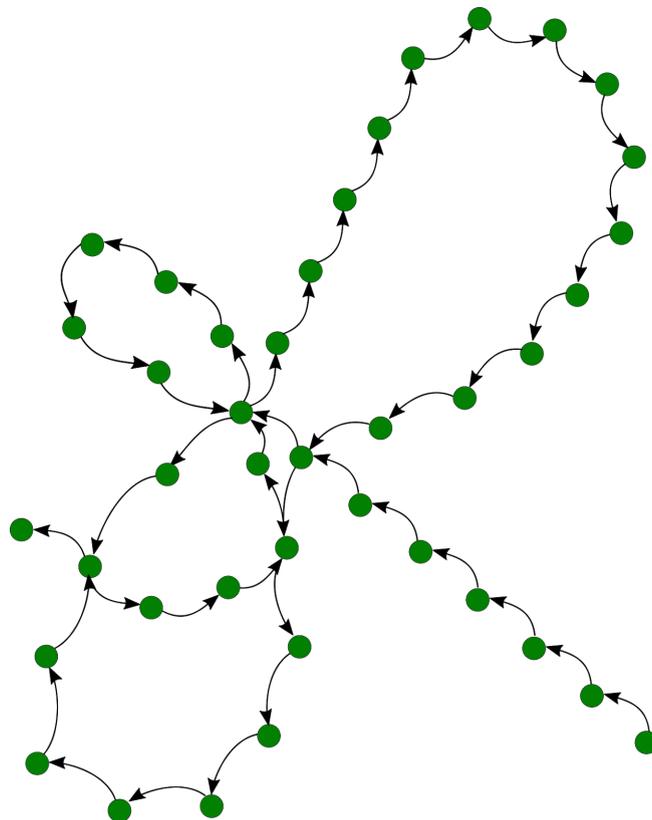


Os grafos são instrumentos matemáticos ricos, mas também são instrumentos importantes, muito importantes, para a linguística e para as ciências cognitivas, pois permitem estudar funções cognitivas diversas na vigília e no sono.

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!



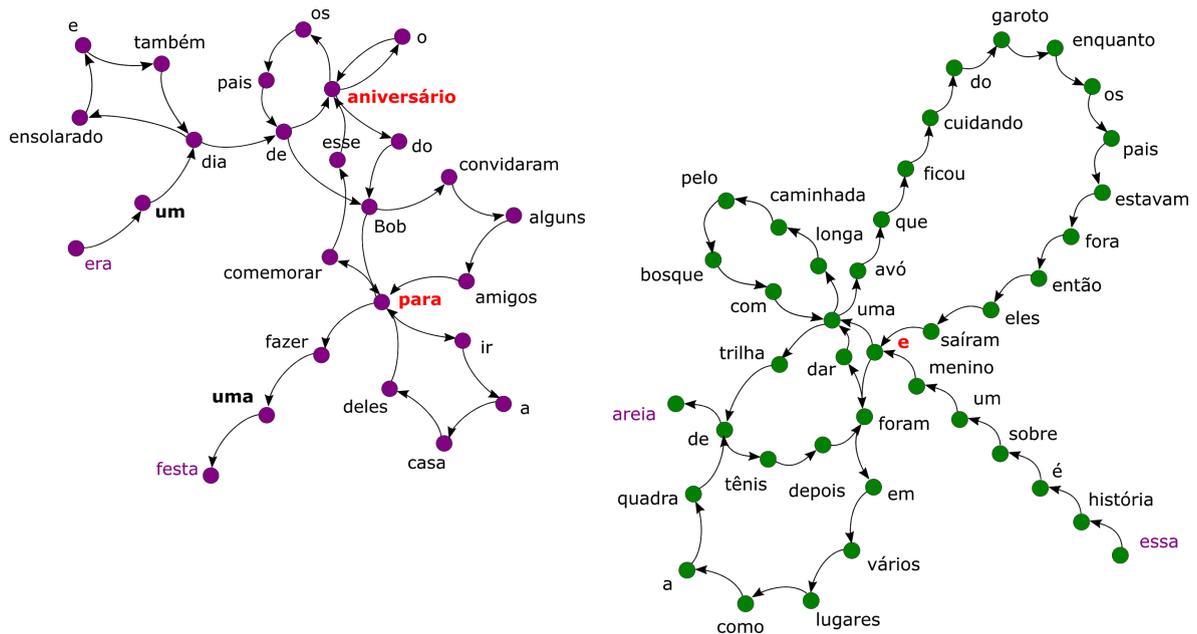
Essa história é sobre um menino e uma avó que ficou cuidando do garoto enquanto os pais estavam fora então eles saíram e foram em vários lugares como a quadra de tênis depois foram dar uma longa caminhada pelo bosque com uma trilha de areia

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

Resposta: E



Temos acima os grafos que representam as falas das duas crianças, dessa vez com as palavras marcadas em cada vértice. Com isso, vamos analisar as alternativas:

(a) De fato, as palavras "aniversário" e "para" têm grau 6, pois existem 6 arestas ligadas a cada uma delas (chegando ou saindo). Isso também poderia ser visto pelo texto: ambas as palavras aparecem três vezes no texto, sem ser nem no início nem no fim. Todas as outras palavras no



texto aparecem no máximo duas vezes, e seus vértices no grafo têm graus 4, 2 ou 1. Portanto, essa alternativa é verdadeira.

(b) O início do grafo é o vértice que tem uma aresta saindo e nenhuma entrando; o fim do grafo é o vértice que tem uma aresta entrando e nenhuma saindo. A palavra "um" é a segunda palavra da frase, e precisa portanto ser o segundo vértice a partir do caminho direcionado. Da mesma forma, a palavra "uma" é a penúltima palavra da frase, e corresponde ao penúltimo vértice. Assim, essa alternativa também é verdadeira

(c) Pelo grafo 1, notamos que o par de vértices "o" e "aniversário" estão ligados por arestas em direções opostas, o que corresponde à definição de arestas paralelas (PE). As arestas paralelas permitem que se vá de do vértice "o" para "aniversário" e vice-versa (em outras palavras, formam um ciclo). Assim, eles formam um componente conexo, e também um componente fortemente conexo. Essa alternativa é, portanto, verdadeira.

(d) O maior componente fortemente conexo (LSC) do grafo 2 é o número de vértices que compõem o maior ciclo do grafo, e vale portanto 33 – que é a quantidade de palavras **distintas** do trecho "[...] **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 [...]**". Essa alternativa também é verdadeira.

(e) No grafo 2, a primeira palavra com grau maior do que 2 é a palavra "e", que tem grau igual a 4 e aparece logo antes da palavra "uma" na frase. Assim, essa é a alternativa incorreta.

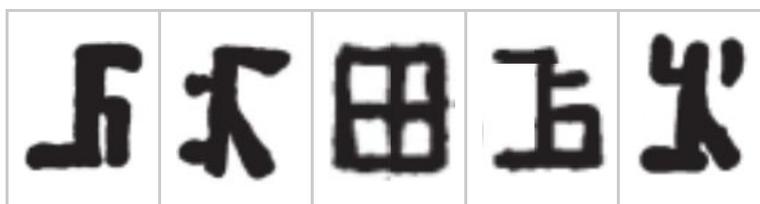
Para saber mais, o material deste problema foi extraído principalmente do seguinte artigo:

MOTA, N.B.; WEISSHEIMER, J.; MADRUGA, B.; ADAMY, N.; BUNGE, S.A.; COPELLI, M.; RIBEIRO, S. (2016) A Naturalistic Assessment of the Organization of Children's Memories Predicts Cognitive Functioning and Reading Ability. *Mind, Brain and Education* (Print), v. 10, p. 184-195.

Os grafos são feitos em um programa chamado SpeechGraphs, desenvolvido no Instituto do Cérebro da UFRN. O software pode ser baixado gratuitamente em: <https://neuro.ufrn.br/software/speechgraphs?id=speechgraphs&controller=softwares>

#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-árabicos, 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

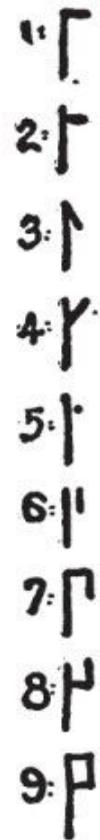
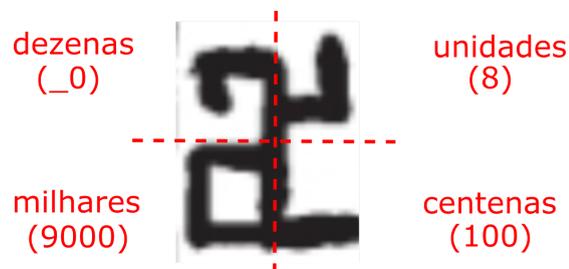


Resposta: D

O sistema de numerais cistercienses é útil para representar números grandes de uma maneira compacta. Neste caso, não seria razoável supor que se trata de uma base de numeração pouco usual, afinal o sistema era usado por monges europeus, cercados por línguas que usam base 10.

Visualmente, é fácil perceber que as figuras são divididas em quatro “quadrantes”, cada um representando unidades, dezenas, centenas e milhares. Resta saber qual é qual.

Uma maneira de começar é perceber que três dos cinco números começam com 1, o que só dá a opção do traço no quadrante inferior esquerdo para representar a casa dos milhares. Além disso, o símbolo do meio, com quatro quadrantes iguais, só pode ser o 9999. Com isso, podemos encontrar o 9 em 1912, concluindo que a casa das centenas é representada no canto inferior direito. Os dois números com repetição entre milhares e dezenas, 1912 e 2421 indicam que as dezenas ficam no canto superior esquerdo e, por fim, os dois números terminando em 1, 1801 e 2421, indicam que o canto superior direito fica com as unidades. Em resumo:



A única alternativa compatível com esses valores é 9178.

Vale observar que os símbolos utilizados no problema são reproduções dos glifos como anotados em dois livros históricos:

- D. C. Beard, *The American boys' book of signs, signals and symbols* (1918).
- L. W. de Laurence, *The great book of magical art, Hindu magic and East Indian occultism* (1912).

Para saber mais, recomendamos a seguinte leitura:

- David King, *The Ciphers of the Monks –A forgotten number notation of the Middle Ages*. Stuttgart: Steiner, 2001. Resumo disponível em <http://www.davidaking.org/Ciphers.htm>

#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>

ㄟㄨˊ, ㄇㄨˋ, ㄌㄞˋ, ㄉㄨˋ, ㄇㄟˊ, ㄎㄟˊ

Quais são as três primeiras palavras faladas no áudio?

- a) ㄌㄞˋ, ㄎㄟˊ, ㄇㄨˋ
- b) ㄟㄨˊ, ㄉㄨˋ, ㄇㄟˊ
- c) ㄟㄨˊ, ㄎㄟˊ, ㄇㄨˋ
- d) ㄌㄞˋ, ㄉㄨˋ, ㄇㄨˋ
- e) ㄌㄞˋ, ㄎㄟˊ, ㄇㄟˊ

— João Henrique Fontes

Resposta: A

As palavras são escritas com dois, três ou quatro símbolos, o último sendo um símbolo pequeno, semelhante a um acento. Uma maneira de começar é notar que, no áudio, há duas palavras que começam com a mesma consoante /m/, que deve ser, portanto, ㄇ. Assim, a primeira letra é a consoante inicial da sílaba.

As duas palavras com /m/ tem três glifos e soam como “mai” e “mei”; uma hipótese que poderia ser feita é que o acento representa o -i final, mas os acentos nas duas são diferentes. Com isso, somos obrigados a concluir que o segundo glifo de ambas representa o ditongo completo, ‘ai’ e ‘ei’ (ㄨ e ㄟ, em alguma ordem).

Todas as sílabas tem a forma consoante-vogais (CV(V)) – exceto duas, que tem um final nasalizado (-ãng); elas devem ser ㄉㄨˋ e ㄎㄟˊ. Assim, faz sentido supor que ㄨ represente esse final nasal.

Em ㄎㄟˊ, podemos notar que o som vocálico deve ter uma correspondente “consonantal” no início da sílaba ㄟㄨˊ. As opções compatíveis com isso, são /i-ou/ e /l-i-ang/. Com isso, descobrimos que ㄉㄨˋ é /h-u-ang/ e ㄌㄞˋ deve ser /ü-e/.

Por fim, era importante notar que o símbolo que parece um acento representa os tons das palavras. /i-ou/ e /ü-e/ são falados sem mudança de tom (tom plano), e por isso não têm acento; /l-i-ang-ˇ/ e /h-u-ang-ˇ/ são falados com um tom que desce e sobe (ˇ); /m-ai-ˊ/ é falado com tom descendente (ˊ), enquanto /m-ei-ˋ/ tem tom ascendente (ˋ), como fazemos em português no final de uma pergunta. Com isso, a resposta correta é A.



#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 umuřnii duk ga alummaa

b) Ɗan 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^wařai

— João Henrique Fontes

Resposta: A

O primeiro ponto do problema é fazer a separação das sílabas. A nota fonética já diz que existem vogais curtas e longas e que as últimas são representadas por vogais duplas. As consoantes duplas, por outro lado, ficariam separadas em sílabas diferentes – como aliás fazemos em português. Com isso, podemos nos assegurar que todos os versos possuíam 12 sílabas métricas, exceto o primeiro, com 13 sílabas.

Como foi dito no enunciado, o ritmo é muito importante na poesia hauçá. Assim, em termos da duração das sílabas, poderíamos separá-las em três tipos: sílabas terminadas em vogal curta (verde), em vogal longa (vermelho), e em consoante ou em i (laranja). Com isso, temos:

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
Al	laa	hu	yaa	yi	u	muř	nii	duk	ga	a	lum	maa
Ɗan	jaa	hi	lii	dab	baa	fa	nee	tun	bay	mu	tu	
Wal	laa	hi	in	har	mun	tša	yaa	mun	bin	ci	kaa	
Zaa	mui	tu	naa	nii	mai	ya	waa	gun	zuu	ci	yaa	
Doo	min	a	loo	tan	nan	da	jaa	hil	cii	k ^w a	řai	

Com isso, fica fácil notar que as sílabas com vogais longas e terminadas em consoante (fechadas) são tratadas como sendo da mesma classe (que podemos chamar de ‘longas’ ou ‘pesadas’), em contraste com as ‘curtas’ ou ‘leves’ que, no padrão dominante, aparecem sempre nas posições 3, 7 e 11. Portanto, o verso que não se encaixa no padrão é o da alternativa A.



Para saber mais, recomendamos os seguintes artigos:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.568.6954&rep=rep1&type=pdf>

https://www.jstor.org/stable/pdf/594988.pdf?casa_token=in2_aFSCd_OAAAAA:uq8Y1AOIozoGfPl4wzc47BOW-sb3H2FTZbymObbRoy4CrpUO2N1oyky9GJGE.m5dL1dvgOQXXm9biM8gtXuq6t1bJmDnlcYFqUfx4BEie3EAVWRHGBKE

#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>	igetán	<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: **ḏ** 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 preenchem as lacunas?

- a) tihwalt, tabaraṭan, initan, iharaḏ
- b) tahwal, tabaraṭen, initan, aharaḏ
- c) tahwalt, tibararaṭan, iniren, iharaḏ
- d) tahwalt, tibararaṭen, initan, aharaḏ
- e) tahwalt, tibararaṭen, inien, aharaḏ

— Rafael Righetto

Resposta: D

As palavras no singular podem ser divididas em dois grupos: as que começam e terminam com **t** e as que começam em vogal, **i** ou **a**. Essa classificação corresponde aos gêneros feminino e masculino em tamahaq, respectivamente – fato que poderia ser suposto a partir do par **abaraḏ/tabaraṭ** ('menino/menina') que, no entanto, não é necessário para resolver o problema.

O fenômeno geral que se percebe é a substituição do **a** na sílaba inicial no singular por um **i** no seu respectivo plural, o que vale para todas as palavras.

Das palavras do gênero masculino, pode-se perceber dois padrões de formação do plural, além de **a > i**:

- palavras terminadas em consoante recebem **-en** (abaraḏ/ibaraḏen; amawey/imaweyen)
- palavras terminadas em vogal recebem **-tan** (ige/igetani; abaraqqa/ibaraqqatan)

As palavras de gênero feminino (que estão entre **t-** e **-t**) recebem o sufixo **-en** para a formação do plural. Este sufixo é adicionado à última consoante da palavra (sem contar **-t**), como em **t-urif-t/t-urif-en** e **t-amar-t/t-imar-en**. Na palavra **tusit**, como não havia consoante antes de **-t**, seu plural se torna **tisuten**. Portanto, a alternativa correta é a D.



#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) גושום, גשוף

c) זיפור, גשוף

d) זיפור, גשון

e) גשום, ציפור

— Jade Yarden Steinmetz, Rafael Righetto



Resposta: C

Ashkenazim e *sefaradim* são os nomes dados a dois grandes grupos de judeus que remontam, respectivamente, à Europa Central e à Espanha medievais.

O problema apresentava todas as palavras em duas formas caligráficas distintas; com isso, era importante saber distinguir que características são diferença de estilo e quais são traços constitutivos das letras. Consideramos que as associações entre os símbolos das duas formas cursivas era relativamente simples; daqui para frente, comentaremos sobre o sistema de escrita usando a ortografia oficial do hebraico.

O primeiro aspecto importante é reconhecer que ambas as formas cursivas são escritas da direita para a esquerda. Isso poderia ser percebido, por exemplo, com a primeira e a quarta palavras, ambas terminando em **n** (נ). Da mesma maneira, **u** (ו) é o segundo símbolo da direita para a esquerda em **shutafah** (שׁוֹתָפָה) e o quarto em **pitsuyim** (פִּיִּצְיִים). É possível, ainda, notar a letra **h** (ה), segunda letra em **lehaarich** (לְהַעֲרִיךְ) e última em **shutafah** (שׁוֹתָפָה).

Além disso, em hebraico, como em árabe, a escrita das consoantes é enfatizada; apenas algumas vogais, como **u** (ו) e **i** (י), são escritas. Isso poderia ser notado, por exemplo, comparando-se **pitsuyim**, com sete símbolos tanto em hebraico quanto em latim, e **shutafah**, onde há cinco símbolos no hebraico, mas sete em latim. Da mesma forma, **menagen** (מִנְגֵּן), com quatro símbolos em hebraico, tem apenas as consoantes escritas.

Um detalhe delicado aqui é a letra **ayin** (ע) que tem um som específico (fricativa faringal); em transcrições latinas, ela às vezes é representada como ‘ mas, nesta transcrição, não é representada. Assim, ela poderia ser confundida com um **a**, mas isso não seria consistente com a maioria das letras **a** do problema. A letra **ayin** aparece em **leha(‘)arich** (לְהַעֲרִיךְ) e em (‘)achshav (עַכְשָׁו). Esse detalhe, de todo modo, não interfere na resolução do problema.

Em hebraico, a letra **vav** (ו) pode representar tanto o som de **u**, como **v** em **achshav** (עַכְשָׁו) e **o** em **tsipor** (צִיפּוֹר), o que também não prejudicaria a resolução do problema. Algumas letras possuem um formato diferente caso ocorram no final, recebendo o adjetivo **sofit** (סּוֹפִית), de **sof** (‘final’). No problema isso ocorre com as letras **mem** (מ), **nun** (נ) e **chaf** (כ), cujas versões **sofit** são ם, ן, e ך, respectivamente.

Com tudo isso, é possível descobrir as formas das palavras **gashum** (גָּשׁוּם) e **tsipor** (צִיפּוֹר).



#9 · Mboi-tata

Mboi-tata, ou *Boitatá*, 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 'djangó'; 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

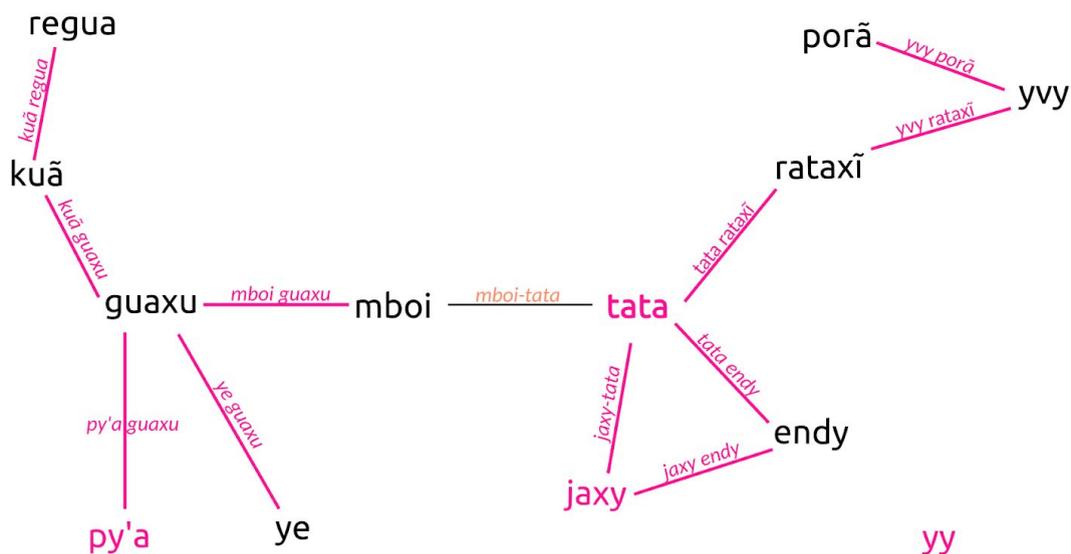
e) yy guaxu, jaxy kuã

— Artur Corrêa Souza



Resposta: A

Em problemas assim, com traduções em ordem aleatória, é melhor organizarmos as palavras de um jeito visualmente compreensível. Podemos juntar as palavras do corpus em um grafo, em que cada traço (ou aresta) corresponde a uma das palavras do corpus, e cada vértice é um dos radicais das palavras: rosa se está contido no corpus e preto se não está.



[1] Uma boa maneira de começar é olhando para a palavra principal do problema: **mboi-tata**. Pelo enunciado, **mboi-tata** significa cobra-de-fogo. Ou seja, **mboi** e **tata** significam cobra e fogo, em alguma ordem. Na tabela, apenas **tata** aparece sozinho e, nas traduções, vemos apenas *fogo*. Portanto, **mboi** = *cobra* e **tata** = *fogo*. Esse primeiro passo facilita bastante os próximos. Temos uma palavra composta relacionada a cobra e três relacionadas a fogo. A palavra com cobra é **mboi guaxu**, que deve significar *sucuri*. Como aprendemos no enunciado, a *sucuri* é a maior cobra do mundo, então **guaxu** deve significar algo como ‘grande’. Inclusive, em algumas versões da lenda do boitatá, a *sucuri* é chamada de ‘boiguaçu’, uma adaptação a partir do nome tupi ou guarani.

[2] As palavras relacionadas a **tata** (fogo) correspondem à parte frontal da onça do grafo. Duas palavras claramente relacionadas a fogo são *fumaça* e *luz do fogo*, mas podem ser qualquer uma das três. Outra palavra relacionada a *luz* é *luar*, que significa a ‘luz da lua’. Então, temos *fogo/luz do fogo* e *lua/luz da lua*. Ou seja, estamos procurando uma palavra para *luz* que apareça junto com **tata** (*luz do fogo*), que apareça junto com uma outra palavra (*luz da lua*) e que esta última esteja no corpus (*lua*). A única possível é **endy**:

tata endy	jaxy endy	jaxy
<i>luz do fogo</i>	<i>luar (luz da lua)</i>	<i>lua</i>

Então **jaxy-tata**, a outra relacionada a lua, só pode ser *estrela* (literalmente lua-de-fogo) e **tata rataxĩ**, *fumaça*. Por enquanto, **rataxĩ** pode significar ou algo do tipo ‘pó’ ou ‘produto’. Até aqui, já fizemos a correspondência correta de 7 das 15 palavras do problema.



[3] Da parte frontal da onça, restam duas palavras da cabeça:

yvy rataxĩ ~ **yvy porã**

Temos uma palavra que é a ‘fumaça’ de **yvy**, além de uma derivada, **yvy porã**. Das palavras restantes, a mais similar à fumaça é *poeira*. Isso nos leva a pensar que **rataxĩ** significa mesmo algo como ‘pó’. Poeira pode ser feita de várias coisas como pele morta, terra, areia, etc. A fonte de poeira com uma palavra derivada é *terra*, com *terra boa*. Então, **yvy rataxĩ** é *poeira* (pó de terra), e **yvy porã**, *terra boa*.

[4] Na parte de trás da onça, temos três palavras para coisas grandes.

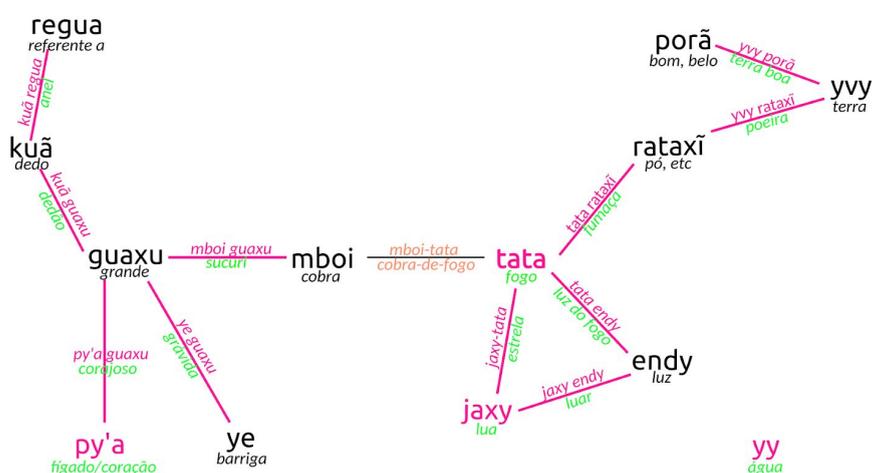
kuã guaxu ~ **kuã regua**
py'a guaxu ~ **py'a**
ye guaxu

As palavras restantes são *água*, *anel*, *corajoso*, *dedão*, *fígado/coração*, *grávida*. Dentre elas, a que tem referência mais clara a algo grande é *dedão* (referindo-se ao polegar). Dedo, por sua vez, deve estar relacionado a *anel* (‘joia do dedo’ ou ‘acessório do dedo’). Como *dedão* e *anel* são ambas compostas, só podem ser **kuã guaxu** e **kuã regua**. Na verdade, **regua** significa, entre outras coisas, ‘algo referente a’, então *anel* é ‘algo referente ao dedo’.

Das últimas quatro, o único par relacionado é *corajoso* (relacionado a sentimento, emoção) e *fígado/coração*. Então, devem ser **py'a guaxu** e **py'a**, o que nos indica que *corajoso* é literalmente ‘coração grande’. Importante perceber aqui que **py'a** pode significar fisicamente o fígado, o coração ou até o estômago, mas significa também a ‘fonte das emoções’, papel feito pelo coração na nossa cultura.

Sobram então *água* e *grávida*. O mais razoável é que *água* seja uma palavra simples, **yy**, e *grávida* seja algo relacionado a grande, **ye guaxu**, que deve ser ‘barriga grande’.

O nosso grafo com as traduções fica então:



Para responder à tarefa, devemos relacionar as palavras que aprendemos no problema. *Tranquilo* deve ter uma formação parecida com *corajoso*, ou seja, ‘coração + algo’. Das palavras que temos, a que mais faz sentido para isso é **porã** (bom, belo). Então, *tranquilo* é **py’a porã** (coração bom). *Nevoeiro* deve ser como *fumaça* e *poeira*, uma formação com **rataxĩ**. O nevoeiro é feito de partículas de água, então só pode ser **yy rataxĩ**, “pó de água”. A correta é portanto a alternativa (a).

Para saber mais:

- História e cultura guarani
<https://historiaeculturaguarani.org/>
 - Lenda do mboi-tata, por Simões Lopes Neto
<http://www.dominiopublico.gov.br/download/texto/bv000122.pdf>
-



#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

Resposta: B

Diferente dos problemas mais comuns de comparação entre línguas, aqui não esperávamos que o participante apenas notasse as transformações regulares, mas soubesse avaliar o que é razoável de se deduzir dos dados, em termos de línguas que não existem mais. Vamos analisar cada uma das alternativas.

1. O alongamento da vogal no japonês é resultado de pelo menos dois processos diferentes: do apagamento da consoante final **-ng** (em ‘professor’, **sheng/sěng/sē**; em ‘masculino’, **xing/sǒng/sē**) quanto da desditongação de algumas sílabas (em ‘quinto’, **hao/ho/gō**; em ‘onda cerebral’, **nao/no/nō**). Outros exemplos além do corpus são *ren.kou/in.ku/jin.kō* (população) e *yin.hang/eun.hěng/gin.kō* (banco).

2. Há dois padrões de correspondência para o **x** do mandarim: **x-s-s** (em ‘professor’, **xien/sǒn/sen**; em ‘notícia’, **xin/sin/sin**) e **x-h-k** (em ‘perigo’, **xien/hǒm/ken**). O **x** do mandarim, assim como outras consoantes como o **j** (**tx**) e o **q** (**txh**), deriva de processos de palatalização de consoantes, por isso só aparece antes de vogais frontais, **i** e **ü**.

3. Há também dois padrões de correspondência para o **w** do mandarim: **w-m-b** (em ‘cultura’ e em ‘notícia’, **wun/mun/bun**) e **w-vogal-k/g** (em ‘perigo’, **wei/ui/ki**; em ‘quinto’, **wu/o/go**). O entendimento acadêmico é que os dois padrões derivam respectivamente das iniciais **mj-** e **ng-** do CM. Em ambos os casos, no japonês as iniciais nasais se tornaram iniciais plosivas sonoras, **b-** e **g-**.

4. Ao analisar as correspondências com o **h** no japonês, podemos encontrar **p-p-h** (em ‘oito anos’, **pa/pal/hati**; em ‘onda cerebral’, **po/po/ha**) e a versão ligeiramente diferente **f-p-h** em ‘invenção’: **fa/pal/hatsu**. Mas esta última não deveria ser considerada uma origem separada, já que temos o mesmo **p** em coreano, nos três exemplos; além disso, o som **/f/** é bastante próximo a **/p/**: ambas articuladas nos lábios, mas uma fricativa e outra oclusiva. De fato, o **f** do mandarim é resultado de uma inicial **pj-** no CM.

Portanto, estão corretas apenas 2 e 4.

Para saber mais:

https://en.wikipedia.org/wiki/Sino-Xenic_pronunciations
https://en.wikipedia.org/wiki/Historical_Chinese_phonology





#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 pĩrôrã 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 pĩrôrã 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'. † 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'âsã
- 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'âsĩ
- d) masí mi'i yaá pekâ direáwĩ; mi'i akaweré yi'î wa'îrã ba'âsĩ
- e) masí mi'i pekâ direáwã; mi'i akaweré yi'î wa'îrã ba'âisã

— Roger Antunes, João Henrique Fontes

Resposta: D

Uma vez que o vocabulário básico seja identificado, é possível identificar alguns elementos básicos, como a ordem das frases Sujeito - Objeto - Verbo (SOV), e a marca de plural -rã. Depois disso, é possível passar aos fenômenos principais do problema.

Um deles se refere aos dois tipos de *relação de posse*. Uma delas é formada só pela justaposição dos termos, como em **mi'i pakí** (você-pai, 'seu pai'), mas também com parente, cachorro, cobra, rato, mãe, pai. A outra fé formada com a partícula **yaá**, como em **yi'î yaá kométi** (eu-yaá-panela, 'minha panela'), que também vale para abacaxi e canoa. Com isso, podemos perceber que o primeiro modelo se aplica sempre a seres animados (pessoas, animais) e o segundo, a objetos inanimados.

O segundo fenômeno principal diz respeito à estrutura 'Eu vi/percebi que...'. Comparando as frases, é fácil perceber que essa ideia é expressa com sufixos no final do verbo, de maneira similar ao que foi explorado na língua tuyuka, no problema 5 da fase 2 da edição Yora. De fato, trata-se de *evidenciais*, ou seja, marcas sobre como a pessoa que fala aprendeu sobre a ação. Com isso, podemos montar a seguinte tabela:

	passado imediato (‘acabou de fazer’)	passado remoto (‘fez’)	não-passado (‘faz’)
eu vi que...	-ami/-ama	-?/-wã	-mi/-?
eu percebi que...	-?/-asã	-?/-?	-sami/-sama

A partir dela, podemos entender que a marca singular/plural é dada pela alternância entre **i** e **a**. Com tudo isso, as estruturas das frases pedidas serão algo da forma:

[Homem] [Você **yaá** lenha] [cortar-**wĩ**]

[Você parente] [Eu peixe] [comer-**asĩ**]

#12 · Citos Gadījumos

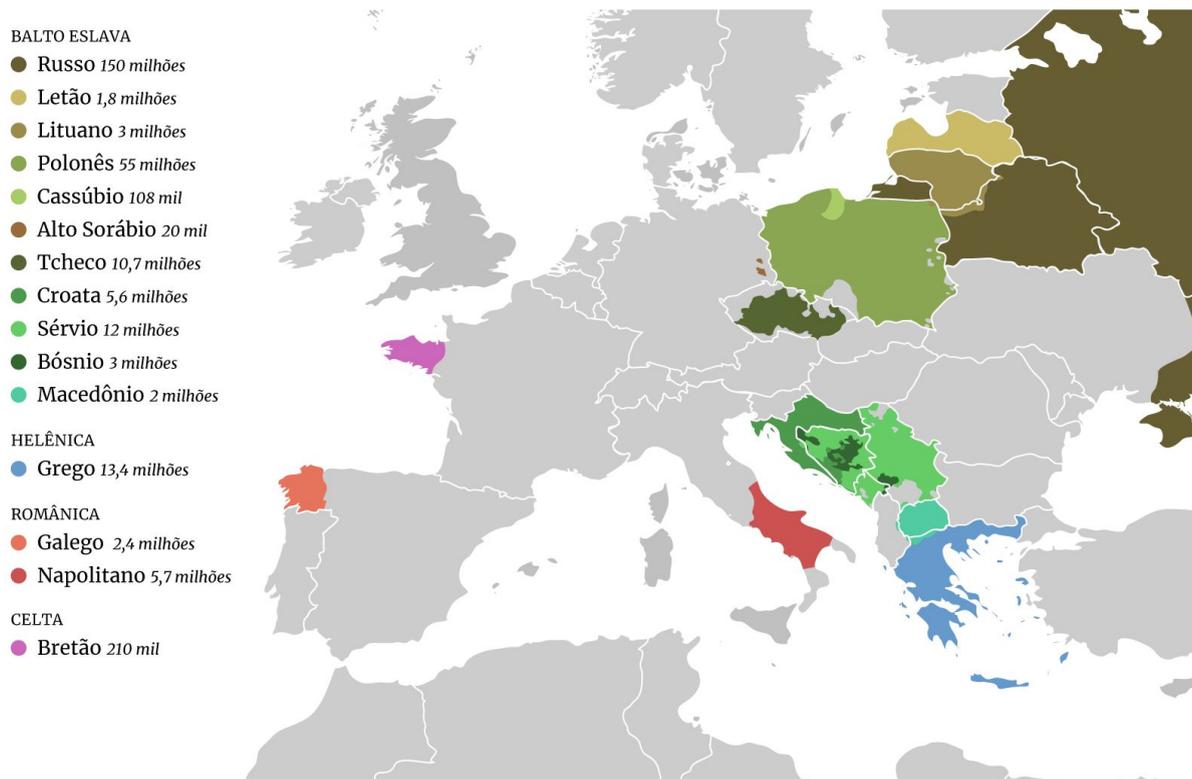
Seguem os nomes dos primeiros quatro meses do ano em várias línguas europeias:

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) sečn+k, choiva, marc, apriel

c) sečn+k, febreiro, strëmiannik, kwiecień

d) sečn+k, choiva, strëmiannik, apriel

e) januari, febreiro, strëmiannik, kwiecień



Resposta: E

Há dois padrões de nomes dos meses no problema: meses que são derivados dos nomes romanos (que é também o caso dos nomes para os meses em português e em inglês), e nomes tradicionais, pré-romanos.

Napolitano, Bretão e Galego são faladas em territórios que faziam parte do lado ocidental do Império Romano e que, portanto, se tornaram territórios principalmente católicos. Todos eles utilizam os nomes latinos para os meses.

Na Europa Oriental (que, no caso deste problema, envolve as línguas balto-eslavas e o grego) temos uma situação menos clara. Um ponto crítico são croata, bósnio e sérvio, línguas muito próximas (muitos linguistas consideram-na uma única língua, o servo-croata) e faladas em territórios bastante imiscuidos; croata, entretanto, usa nomes eslavos para os meses, mas bósnio e sérvio usam nomes romanos. Dentre os dados fornecidos, a maior diferença entre seus territórios é a religião: enquanto a Croácia é essencialmente católica, a Sérvia e a Bósnia são territórios ligados à Igreja Ortodoxa (que tradicionalmente possui uma centralidade na Grécia). A Bósnia, na verdade, é um país de maioria muçulmana (por influência posterior do Império Otomano), mas isso não afetou o nome dos meses na língua bósnia.

Estendendo esse raciocínio para os demais países orientais, vemos que esse padrão se aplica: territórios balto-eslavos de maioria católica mantêm os nomes eslavos para os meses, enquanto os países balto-eslavos de maioria ortodoxa (e também a Grécia), provavelmente por influência histórica da Igreja Ortodoxa, usam os nomes romanos.

Assim, em macedônio, de maioria ortodoxa, deve se usar **januar** e não *sečn ik*; em cassúbio e em polonês, de maioria católica, deve se usar **strěmiannik** e **kwiecień**, não *marc* nem *apriel*. Portanto, a alternativa correta é a E.

É interessante notar que a separação entre os sistemas não é completa. Em polonês, por exemplo, embora a maioria dos nomes tem origem eslava, pelo menos dois meses – *marzec* e *mai* – tem nome de origem latina. Além disso, mesmo nos países que usam oficialmente os nomes romanos, os antigos nomes eslavos ainda são conhecidos por algumas pessoas, e eventualmente usados em contextos bastante específicos.

Os calendários eslavos tradicionais normalmente usam nomes associados ao ciclo da natureza e dos trabalhos agrícolas. Por exemplo, *březen* (março), em tcheco, possivelmente vem de *březí*, ‘grávida’, uma alusão à vida renascendo no início da primavera. Também em tcheco, *květen* (maio) vem de *květ*, ‘florescer’; em polonês, entretanto, a mesma metáfora é usada pra outro mês: *kwiecień* (abril). Em croata, *kolovoz* (agosto) é o mês de *kolo voziti*, ‘dirigir a carroça’; o mesmo mês, em macedônio, é chamado de *žitar*, o mês dos cereais, enquanto o mês anterior, julho, é chamado de *zlatec*, o mês dourado.

Por fim, é importante mencionar que, mesmo em territórios que usam apenas os nomes romanos, alcunhas populares para os meses sempre surgem, associadas à agricultura, festividades especiais etc. É o caso, por exemplo, em galego, em que é possível encontrar falantes se referindo a *xuño* como *San Xoán* ou a *xullo* como *mes da seitura* (colheita).

