(D) Let's Pivot! (1/3) [10 Points]

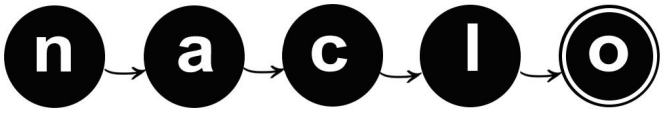
Languages and their speakers have many ways of describing colors. English has 11 "basic color terms": *white*, *gray*, *black*, *red*, *orange*, *yellow*, *green*, *blue*, *purple*, *pink*, and *brown*. However, other languages can have as few as two or as many as 12.

Below is a table showing eight objects and the basic color word that native speakers of 15 different languages might use to describe them.

	ripe plum	*ripe strawberry*	*ripe banana*	*fresh grass*	*clear sky*	*deep ocean*	*adult crow*	*fresh snow*
Bassa	**hui**	**ziza**	**ziza**	**hui**	**hui**	**hui**	**hui**	**ziza**
Bété	**kp**	**zeli**	**zeli**	**kp**	**kp**	**kp**	**kp**	**fee**
Ejagham	**ebi**	**ebi**	**ebi**	**enyaga**	**enyaga**	**enyaga**	**enyaga**	**ebare**
Karajá	**iso**	**iso**	**tyre**	**tyre**	**tyre**	**tyre**	**ilyby**	**ura**
Matsés	**piu**	**piu**	**piu**	**umu**	**umu**	**umu**	**chëshe**	**ushu**
Apinaye	**grãgrã**	**kamrẽk**	**ràràr**	**grãgrã**	**grãgrã**	**grãgrã**	**tyk**	**aka**
Tsafiki	**luban**	**luban**	**laqueban**	**losimban**	**fiban**	**paban**	**paban**	**fiban**
Seri	**kʔe'el**	**kʔe'el**	**kwassool**	**ko'il**	**ko'il**	**ko'il**	**k'ooppool**	**ko'oxp**
Guambiano	**piguig**	**piguig**	**uscuig**	**chilga**	**pillig**	**pillig**	**yalig**	**polig**
Teribe	**dindin**	**sresren**	**shoylor**	**keson**	**dindin**	**dindin**	**sisi**	**pluplun**
Tlapanec	**minuu**	**mana**	**mojmo**	**maxa**	**maxa**	**minuu**	**skuni**	**mixa**
English	**purple**	**red**	**yellow**	**green**	**blue**	**blue**	**black**	**white**
Yakan	**taluk**	**peat**	**binaning**	**gaddung**	**bilu**	**bilu**	**ittem**	**pole**
Amuzgo	**tsjan'chi**	**wee**	**cajan**	**cachuii'**	**tsa**	**tsjo'**	**ntom**	**canchii'**
Hebrew	**sagol**	**adom**	**tsahov**	**yarok**	**t'khelet**	**kakhol**	**shakhor**	**lavan**

Note that this table of color terms also appears in problem (C) from NACLO 2025, Round 1. Otherwise, problems (C) and (D) of NACLO 2025, Round 1 are not related — you don't need to solve either one in order to solve the other.

Information about these languages can be found on page (3/3). The language information is not relevant for solving this problem.



(D) Let's Pivot! (2/3)

Imagine that you know a way to translate reliably from language *A* to language *B*, and also a way to translate from *B* to language *C*. Then you also know how to translate from *A* to *C*: just translate from *A* to *B* first, and next from *B* to *C*. When a machine translation system uses this strategy, *B* is called a *pivot language*.[1]

NACLO Labs is about to release its groundbreaking new app, *NacloTranslate*. *NacloTranslate* can translate between more language pairs than any of its competitors, and it does so by using pivot languages.

Unfortunately, *NacloTranslate* doesn't always translate totally reliably. In a study, NACLO Labs scientists tried translating **ebi** from Ejagham to Seri, using Karajá as a pivot language. They hoped to see the app give only the answers **kʔe'el** and **kwassool**, which are the results that could be produced from direct (no pivot) translation. But it was observed producing each of these three results on different test runs:

Ejagham ⇄ Seri	
ebi	**kʔe'el**

Ejagham ⇄ Seri	
ebi	**kwassool**

Ejagham ⇄ Seri	
ebi	**ko'il**

D1. Below are three more *NacloTranslate* test runs, from a different study. A pivot language was used for exactly one of the three language pairs—which one? Give a possibility for what the pivot language could be.

A.

Tlapanec ⇄ Matsés	
minuu	**umu**

B.

Yakan ⇄ Guambiano	
ittem	**chilga**

C.

Hebrew ⇄ Teribe	
kakhol	**dindin**

Test run that used a pivot (A, B, or C): ☐

Give one possible pivot language, from the 15 options on page (1/3): ☐

D2. NACLO Labs is considering its options for having *NacloTranslate* translate **paban** from Tsafiki to English.

a. What are the results of a direct translation? Select <u>all possible</u> answers. Make your selections by filling in the box just to the left of the word:

☐ **purple** ☐ **red** ☐ **yellow** ☐ **green** ☐ **blue** ☐ **black** ☐ **white**

b. What are the results if Karajá is used as a pivot? Select <u>all possible</u> answers. Make your selections by filling in the box just to the left of the word:

☐ **purple** ☐ **red** ☐ **yellow** ☐ **green** ☐ **blue** ☐ **black** ☐ **white**

c. NACLO Labs selects a different pivot language, and the results for **paban** are **green**, **blue**, and **black** (and no others). Which languages could NACLO Labs be using as pivot? Write <u>all three</u> possible language names, in any order, choosing from the 15 options on page (1/3):

☐

[1] This strategy can be useful when training data is hard to find. Modern machine translators may need to be trained on huge amounts of *parallel text* — that is, text in one language given alongside its translation in another (much like in some NACLO problems!). Sometimes, *A-C* parallel text will be hard to find, but *A-B* and *B-C* parallel text will be much more common.

(D) Let's Pivot! (3/3)

D3. Here are six test runs of *NacloTranslate*, from a different study. From the options below, identify the pivot language used in each test run by matching options A-F to test runs 1-6. Each option was used once.

1.

Guambiano ⇄ Tlapanec	
piguig	**skuni**

2.

Tlapanec ⇄ Amuzgo	
minuu	**tsa**

3.

Seri ⇄ Yakan	
kʔe'el	**taluk**

4.

Matsés ⇄ Seri	
piu	**ko'il**

5.

Amuzgo ⇄ Matsés	
cajan	**ushu**

6.

Yakan ⇄ Guambiano	
gaddung	**piguig**

Options: A. Apinaye B. Bassa C. Bété D. English E. Teribe F. Tsafiki

Write one letter (A-F) per box to match each language option to the test run in which it was the pivot:

1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐ 6. ☐

Information on the languages featured in this problem is below. Note that this information is not necessary or helpful for solving the problem.

Amuzgo and Tlapanec are clusters of closely related Otomanguean languages spoken in Mexico by around 60,000 and 150,000 respectively. Apinaye and Karajá are Macro-Jê languages spoken in Brazil by around 2000-3000 people each. Bassa and Bété are Kru languages. Bassa is spoken by around 600,000 people in Liberia, Ivory Coast and Sierra Leone. Bété is spoken by several thousand people in Ivory Coast, although the dialect used in this problem had only 50 remaining speakers in 1992. Ejagham is in the Bantoid subgroup of the Atlantic-Congo language family, and is spoken by around 120,000 people in Nigeria and Cameroon. English is an Indo-European language with around 400 million native speakers, and close to 1.5 billion total speakers worldwide. Guambiano and Tsafiki are Barbacoan languages. Guambiano is spoken by around 20,000 people in Colombia. Tsafiki is spoken by around 2000 people in Ecuador. Hebrew is an Afro-Asiatic language from the Semitic subgroup. It was extinct as a spoken language between the 5th and 19th century CE, although it continued to be used for religious purposes; now it has around 5 million native speakers, and around 9 million total speakers, primarily in Israel. Matsés is a Panoan language spoken by around 2000 people on the Peru-Brazil border. Seri is a linguistic isolate spoken by 700-1000 people in Mexico. Teribe is a Chibchan language spoken by around 3000 people in Panama and Costa Rica. Yakan is an Austronesian language, spoken by around 100,000 people in the Philippines.