(C) Color Me Confused (1/4) [15 Points]

Languages and their speakers have many ways of describing colors — sometimes by comparing to well-known materials (copper hair, mustard wallpaper); sometimes using very specific terms (crimson ribbon); and sometimes using very broad terms (gray-blue sky). 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	k?e'el	k?e'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 (D) 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, and others that appear in this problem, can be found on page (4/4). Note that the language information is not relevant for solving this problem.



(C) Color Me Confused (2/4)

In 1969, the anthropologist Brent Berlin and linguist Paul Kay proposed a set of rules about basic color terms. Five of them are given below in a simplified form, with the objects (*ripe plum, ripe strawberry*, etc.) replaced with the letters (A) to (H). Note that a word is described as *unique* if it is used to describe only one of the eight objects.

1.	Every	language	has	at	least	two	basic	color	terms,	which	distinguis	sh
	betwee	en (A) and	d (B)									

- 2. If it has at least three, it uses different words for each of (A), (B) and (C). If it has only three, the same word is used for both (B) and (D).
- 3. If it has at least five, it distinguishes between (B) and (D), and also has a unique word for (E).
- 4. If it has at least six, it distinguishes (D) from (F).
- 5. A language only has a unique word for (G) if it also has one for (H). If it does not have a unique word for (H), (G) will share a word with at least one of (A), (D) or (F).

Although these rules are no longer believed to be completely true for all languages, they do still describe most languages accurately. In particular, they describe all 15 languages in the table on the previous page.

C1. Determine which object each of the letters (A) - (H) refers to. Each of the eight objects corresponds to exactly one letter. Next to each object, write the letter (A-H) that replaces it in rules 1 to 5:

ripe plum:	ripe strawberry:	ripe banana:	fresh grass:	
clear sky:	deep ocean:	adult crow:	fresh snow:	

Below is a similar table for six other languages.

	ripe plum	ripe strawberry	ripe banana	fresh grass	clear sky	deep ocean	adult crow	fresh snow
Kwerba	nokonum	nokonum	kainanesenum	icem	asiram	icem	icem	asiram
Lele	wile	wile	bore	bore	bore	induwa	induwa	bole
Ocaina	tsipi	tsiiho	boora	moxooso	moxooso	moxooso	jutohfu	furaana
Waorani	opa'tawae	opa'tawae	na'amae'ta	we'nedae	we'nedae	we'nedae	wintamo	na'amae'ta
Yucuna	ipirani	querani	jehuani	ipureni	ipureni	ipureni	cameni	jareni
Zapotec	nu-bizat	nu-na'	nu-cach	nu-ca'	nu-ca'	nu-asul	nu-cas	nu-cacy



(C) Color Me Confused (3/4)

C2. For each language in the table at the bottom of the previous page, indicate whether its color words follow rules 1-5 (YES) or not (NO) by filling in the bubble next to the correct answer: **Kwerba** YES NO) YES) NO <u>Lele</u>) YES NO YES Waorani Ocaina () YES) NO Yucuna Zapotec Below is another similar table, but with one word in each row left blank. ripe ripe ripe fresh clear deep adult fresh plum strawbery banana grass sky ocean crow snow Nafaanra nye finge nye a. wa wэ wэ wэ sõa'ro ya'saro ñiro butiro Tucano sõa'ro ewuro b. ya'saro Kalam likañ walin minj-kimemb muk muk mosimb tund c. C3. For each of the blanks a. - c. in the table above, choose all possible words from the list of options that the language might use to refer to that object, while still following rules 1-5. In each case, the options are all the words shown in the table in that language. Make your selections by filling in the box just to the left of the word, remembering to select <u>all</u> possible answers: nye finge a. sõa'ro ewuro ya'saro ñiro butiro b. likañ walin minj-kimemb muk mosimb tund c.

(C) Color Me Confused (4/4)

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

Amuzgo, Tlapanec, and Zapotec are clusters of closely related Otomanguean languages spoken in Mexico by around 60,000, 150,000, and 500,000 people, 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 and Nafaanra are Atlantic-Congo languages. Ejagham is part of the Bantoid subgroup, and is spoken by around 120,000 people in Nigeria and Cameroon. Nafaanra is part of the Senufo subgroup, and is spoken by around 90,000 people, mostly in Ghana. 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 and Lele are Afro-Asiatic languages. Hebrew is from the Semitic subgroup, and 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. Lele is from the East Chadic subgroup, and has around 26,000 speakers in Chad. Kalam is a Trans-New-Guinea language, spoken by around 15,000 people in Papua New Guinea. Kwerba is a Greater Kwerbic language spoken by around 2500 people in West Papua. Matsés is a Panoan language spoken by around 2000 people on the Peru-Brazil border. Ocaina is a Huitotoan language, spoken by around 50 people along the Peru-Colombia 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. <u>Tucano</u> is a Tucanoan language, spoken by around 12,000 people in Brazil and Colombia. Waorani is a linguistic isolate spoken by around 2000 people in Ecuador; more speakers may be living in uncontacted groups in Peru. Yakan is an Austronesian language, spoken by around 100,000 people in the Philippines. Yucuna is an Arawakan language, spoken by nearly 2000 people along the Miritiparaná river in Colombia.

© 2025, Simi Hellsten and Andrew Lamont, North American Computational Linguistics Open Competition, Round 1

