YOUR NAME: REGISTRATION #

(P) Combining Categories in Tok Pisin (1/2) [15 points]

This problem is a follow-up to problem O and has to be solved after that problem. Tok Pisin (also referred to as New Guinea Pidgin or Melanesian Pidgin) is a creole language spoken in the northern mainland of Papua New Guinea and surrounding islands. It is an official language and the mostly widely used language in the country, spoken by over 5 million people.

Many Tok Pisin words come originally from English – its name comes from "talk" and "pidgin"[1] -- but Tok Pisin isn't just English. It has a distinct grammar and uses these words in different (but systematic!) ways.

P1. Below are sentences in Tok Pisin with a scrambled list of English translations. Match each sentence to its English equivalent.

1.	Brata bilong em i stap rit.	E
2.	Ol i stap dringim wara.	H
3.	Ol i ken ritim buk bilong mi.	C
4.	Em i ritim buk pinis.	A
5.	Em i laik rit.	G
6.	Susa bilong em i ken rait.	D
7.	Susa bilong mi i boilim wara.	B
8.	Wara i boil pinis.	F

A.	He has read the book.
B.	My sister boils the water.
C.	They can read my book.
D.	His sister can write.
E.	His brother is reading.
F.	The water has boiled.
G.	He wants to read.
H.	They are drinking water.

P2. Translate the following Tok Pisin sentence into English:

Brata bilong mi i stap ritim buk bilong susa bilong mi.

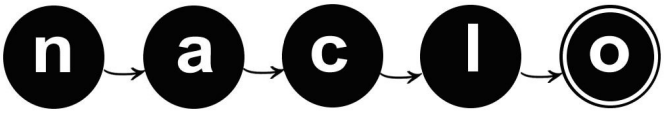
My brother is reading my sister's book.

P3. Translate the following English sentence into Tok Pisin:

Their sister wants to write a book.

Susa bilong ol i laik raitim buk.

[1] A pidgin language is a communicative system developed by two or more groups of people who do not share a common language. Tok Pisin started out as a pidgin but has since developed into a creole, a complex language in its own right.



YOUR NAME: REGISTRATION #

(P) Combining Categories in Tok Pisin (2/2)

P4. Describing these words in terms of their CCG categories (introduced in Problem O) highlights that these aren't English words combined according to English rules, but are Tok Pisin words combined according to Tok Pisin rules.

Match each Tok Pisin word to its CCG category. Some categories will be used more than once. The symbol S_b is short for 'Bare Clause'.

1.	bilong	B
2.	brata	A
3.	boil	D
4.	boilim	E
5.	buk	A
6.	dringim	E
7.	em	A
8.	i	C
9.	ken	G
10.	laik	G

11.	mi	A
12.	ol	A
13.	pinis	F
14.	stap	G
15.	raitim	E
16.	rit	D
17.	ritim	E
18.	susa	A
19.	wara	A

A.	NP
B.	(NP \ NP) / NP
C.	(S \ NP) / (S_b \ NP)
D.	(S_b \ NP)
E.	(S_b \ NP) / NP
F.	(S_b \ NP) \ (S_b \ NP)
G.	(S_b \ NP) / (S_b \ NP)

P5. Explain your answer.

- There is only one way to form S, which is S \ NP (from C)
- A: brata, buk, susa, wara, mi, ol, em can be subjects (and they are to the left)
  so we assume it is NP, which combines with S \ NP to make S.
- B: bilong needs an owner to the right, and forms a new noun phrase (NP \ NP) / NP fits,
  because in buk bilong mi, bilong mi cancels to get (NP \ NP), which cancels with buk to get a NP
- C: analyzing "Em i laik rit", em is NP, so "i laik rit" must be S \ NP. Parsing it as (i, (laik, rit)), (laik,rit) must be (Sb \ NP). Thus assigning C for i gets us S \ NP which is what we needed
- D: "laik rit" must combine to be Sb \ NP. It seems that simple present verbs like rit always appear after "i" or "laik/ken/stap", so we assign D as they combine with "i" and "laik/ken/stap" to form S\NP and Sb \ NP respetively-
- E: "ritim, boilim, raitim" seem to take an object (transitive verb) and succeeds "i", so it must consume NP, and return Sb \ NP. E is (Sb \ NP) / NP.
- F: "pinis" happens after a verb. F works because "i boil pinis" - boil pinis must be (Sb \ Np). Boil is (Sb \ Np) so pinis must consume (Sb \ Np) and return (Sb \ Np). F works.
- G: "laik, ken, stap" happen before verbs. G works - same reasoning as F except G must consume (Sb \ Np) to the right.