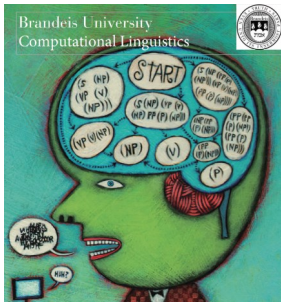


NACLO thanks the following for their generous contributions:



Carnegie Mellon University
Language Technologies Institute



Yale University

SOLUTIONS

The Fifteenth Annual

North American Computational Linguistics Open Competition 2021

www.nacloweb.org

**Open Round
January 28, 2021**

Serious language puzzles that are surprisingly fun!

-Will Shortz, crossword editor of The New York Times and Puzzlemaster for NPR

(A) The Tortoise and the Rabbit (1/1) [Solution]

A1.	Malagasy	English
	(a) Mihaona ny alika ny sokatra.	The tortoise meets the dog.
	Nitsiky ny sokatra mainty.	(b) The black tortoise smiled.

A2.	Malagasy	English
	(a) Miteny ity alika kely ity.	This small dog speaks.
	Niaro ny saka ity bitro lehibe ity.	(b) This large rabbit protected the cat.
	(c) Niarahaba iny sahona kely iny ity vorona maitso ity.	This green bird greeted that small frog.

Vocabulary:

Nouns:	Adjectives:	Determiners:	Verbs:
sokatra = tortoise	kely = small	ny = the	mihaona = meets
alika = dog	mainty = black	ity ... ity = this	miarao = protects
bitro = rabbit	lehibe = large	iny ... iny = that	mitsiky = smiles
vorona = bird	maitso = green		miarahaba = greets
saka = cat			miteny = speaks
sahona = frog			

Morphology:

The present tense of each verb starts with m-. To make it past tense, change the m- to n-.

Word order:

The basic word order of the sentence is Verb-Object-Subject.

Adjectives go after nouns.

“The” goes before the noun.

Demonstratives (“this” and “that”) appear twice, once before and once after the noun. If there is an adjective, it appears inside of the demonstrative pair with the noun (e.g., “this frog green this”).



(B) Junk Mail: Letters We Don't Need (1/2) [Solution]

The alphabet of Duw Ulbubud is summarized in these two tables, which are equivalent. The one on the left is useful for translating from Duw Ulbubud to English while the one on the right is useful for the reverse direction.

Duw Ulbubud letter(s)	English letter(s)
u	a, e, i, o, u, y
b	b, m, p
d	d, n, t
g	c, g, k
l	l
r	r
v	f, v
w	w
z	j, s, z
[nothing]	h (i.e., h is deleted)
gz	x

English letter	Duw Ulbubud letter(s)
a	u
b	b
c	g
d	d
e	u
f	v
g	g
h	[nothing] (i.e., h is deleted)
i	u
j	z
k	g
l	l
m	b
n	d
o	u
p	b
q	g
r	r
s	z
t	d
u	u
v	v
w	w
x	gz
y	u
z	z



(B) Junk Mail: Letters We Don't Need (2/2) [Solution]

Translation of sample texts in Duw Ulbubud (which, by the way, means New Alphabet):

Headline: WORDS WILL BE SPELLED DIFFERENTLY BECAUSE OF "NEW ALPHABET"

Article: For better or for worse, the transformed alphabet will make writing look like a puzzling jumble of random letters. We just have one major question: is it justified? Expect lots of complaints from people who grew up with the old system.

B1. words: wurdz the: du fifth: vuvd squeaky: zgouugu
jazz: zuzz bagpipes: bugbubuz vertex: vurdugz calming: gulbudg

B2. *Moby Dick* in Duw Ulbubud: Bubu Dugg Misinterpretation of this title: Baby Duck

B3. In Duw Ulbubud, *Give me a minute* is written *Guvu bu u bududu*. That last word, *bududu*, is the source of the confusion: even though it's intended to mean *minute*, it could also mean *banana* or *potato*. Thus, the fruit was a banana, and the vegetable was a potato.

B4. Stuffy-Nose Talk

Some Linguistic Background: (This information is totally unnecessary for solving this problem):

Consonants are generally defined by three properties called voicing, place of articulation, and manner of articulation. Voicing describes whether a sound is produced with the vocal folds vibrating or not; if the vocal folds vibrate, the sound is voiced, as is the z sound; otherwise the sound is voiceless, as is the s sound. All consonants are produced by some sort of constriction in the vocal tract, and place of articulation states exactly where the constriction occurs. For example, b and p have a labial place of articulation (i.e., at the lips) because they are produced by closing the lips. Manner of articulation describes how complete the constriction is: some sounds, such as b and p, are called stops and involve a complete closure; others, such as s and z, are called fricatives and involve a nearly complete closure; and still others, such as r and l, are called approximants and do not involve a very large constriction. Consonants can also vary in a property called nasality: sounds produced with the nasal tract open, such as n and m, are called nasal, while other sounds are called oral.

In Duw Ulbubud, vowels are all grouped together, and consonants are put into groups that share place and manner of articulation but that may vary in voicing and/or nasality. The voiced, oral member of the group is the one chosen for inclusion in Duw Ulbubud. (J actually has a slightly different place of articulation than s and z, but it was close enough for those three to be grouped together). X was replaced by two letters because it actually stands for two sounds. H was removed entirely because the letter h is often silent (as in "what" and "hour"), and even in words where h is not silent it often disappears in casual speech (for example, "What did he say?" may actually be pronounced "What did 'e say?").

The problems with Duw Ulbubud mostly arise from the fact that Duw Ulbubud is a phonetically-based system, yet English spelling is not very phonetic. A single English letter may represent multiple sounds; for example, Duw Ulbubud treats y as a vowel, even though it sometimes acts as a consonant, and Duw Ulbubud treats c as if it were always pronounced like k even though sometimes it is produced as the s sound. At the other end of the spectrum, multiple English letters can stand for a single speech sound such as "sh," "th," or "ou," but Duw Ulbubud does not treat these pairs as individual entities but rather as two separate letters. Lastly, English contains many silent letters which generally create strange results when they are transformed into Duw Ulbubud.



(C) Versatile Verb (1/2) [Solution]

C1.

Waama		Literal Translation	
1	Cando dɛbite kpi, o ñ faa o suka.	J	<i>Tchando neighbor died, he and inherited his car.</i>
2	Tando dori.	F	<i>Rain fell.</i>
3	N pe saaki ti yete.	H	<i>My wife swept our house.</i>
4	Bika kɔɔsi kɔɔka.	O	<i>Child sold hen.</i>
5	Soosada kaate.	M	<i>Soldiers assembled.</i>
6	Suka kpi.	N	<i>Car died.</i>
7	Ba kaate tiibu band.	K	<i>They gathered tree under.</i>
8	N yeentire n daaso.	L	<i>I hurt my friend.</i>
9	Bisu yɔkɔɔti.	I	<i>Children had-fun.</i>
10	Tiibu dori puɲa mii.	A	<i>Tree fell forest in.</i>
11	N taka n daaso yete.	C	<i>I went-to my friend house.</i>
12	Maari pei dikitifa, o ñ fa piisi.	E	<i>Marie lost money, she but it found.</i>
13	Suka miiki pɔmpɔmma.	B	<i>Car passed-by earlier.</i>
14	Bika dori.	D	<i>Child fell.</i>
15	N kɔɔka taka Yooto yete.	G	<i>My hen went-to Yooto house.</i>

1. J 2. F 3. H 4. O 5. M 6. N 7. K 8. L
 9. I 10. A 11. C 12. E 13. B 14. D 15.

Notes/Observations about Waama based on the presented data:

There is no possessive marker. The structures *Cando house* and *friend house* work similarly to phrases like *garage door* in English.

Word order is SVO except with object pronouns.

In compound sentences, the coordinating conjunction follows the subject Noun Phrase.

The word *ñ* can mean either *and* or *but* depending on the context.

Definiteness is not marked by means of definite or indefinite articles. Definiteness is determined by context.

Instead of prepositions, the language has postpositions (tree under, forest in).

Lexical observations: *take* means 'go to'; *miiki* means 'pass by'; *kpi* means 'die' in the larger sense of 'come to an end' (see Task 2); *yɔkɔɔti* means 'have fun' as well as 'play' (see Task 3c).

The nouns in this problem become plural as follows: *bika*(sg) – *bisu*(pl).

The third person personal and possessive pronoun *o* is not gender marked (*o* = he/she, his/her).



(C) Versatile Verb (2/2) [Solution]

C2. One of the verbs you encountered in the sentences in the column on the left above would be used by speakers of Waama in all 4 sentences below, each of which is translated into English. Which of the verbs you encountered above is that verb?

Waama	English
Sentence 1 N tokore <u>kpi</u> .	'My shirt is torn.'
Sentence 2 Yaama <u>kpi</u> .	'The matter is settled.'
Sentence 3 O beere <u>kpi</u> .	'He lost his fame.'
Sentence 4 Yima <u>kpi</u> .	'The water is frozen.'

C3. Translate into Waama.

The children gathered under the house.	Bisu kaate yete band.	(Note: 'under' / <i>band</i> is a postposition)
I sold my car.	N kɔɔsi n suka.	
Her friend played in the rain.	O daaso ɣɔkɔɔti tando mii.	(<i>ɣɔkɔɔti</i> = 'have fun = play') (<i>tando dori</i> = 'rain fell') (<i>in/mii</i> is a postposition)

C4. Translate into English.

Ba kɔɔsi ti kɔɔsu.	They sold our hens. (bika-bisu → kɔɔka-kɔɔsu)
N susu kpi.	My cars broke down. (bika-bisu → susa-susu)
Maari daaso fa faa.	Marie's friend inherited it.



(D) Made You Look! (1/3) [Solution]

D1.

	Hypothesis 1	Hypothesis 2	Hypothesis 3	Hypothesis 4	Hypothesis 5
Mary had a little *lamp*	?,E	U	?,U,E	?,U,T,E	(a) U,E
I locked the door with a *steering wheel*	?,W	(b) O	?,W	?,M,B,W	O,W
In the desert sky I could see the *Milky Way*	?,D	D	?,P,K,D	(c) ?,P,K,O,D	D
This fish is a *clamp*	(d) ?,Q	(e) C	(f) ?,X,Q	(g) ?,U,X,T,E,Q	C,X,Q
The floor needs a *carpet*	(h) ?,V	(i) V	(j) ?,L,C,V	(k) ?,L,C,V	(l) V

Here is what the 5 hypotheses are:

Hypothesis 1: Wait until the entire utterance has been completed, then look at the image for word that was uttered. This means that the participant will always start out looking at no image in particular while they wait for the utterance to finish, and then they will look at the image for word that was uttered.

Hypothesis 2: Without listening to the word at all, immediately look at the image for the most likely next word, and never look at anything else. For example, for *I locked the door with a _____*, the most likely word to go in the blank is *key*, so a participant obeying Hypothesis 2 would only look at the key.

Hypothesis 3: Whenever the part of the word that has been uttered so far forms a word, look at the image for that word. For instance, for *In the desert sky I could see the *Milky Way**, the participant first has not heard any complete word at the start, so they don't look at any image in particular; and then they hear the word *mill*, so they look at the mill; then they hear *milk*, so they look at the milk; and finally they hear *Milky Way* and look at the Milky Way.

Hypothesis 4: Whenever you hear the end of a complete word, look at the image for that word. If multiple words are completed at once, start with the shortest one then work to the longest one. For example, for *This fish is a *clamp**, the participant has not heard a complete word at the start, so they start of looking at no complete image. Then, after the *clam-* part of the word has been uttered, that finishes the word *clam* and the word *lamb*, so the participant first looks at *lamb* (the word with fewer sounds) followed by *clam*. Then, when the full *clamp* has been uttered, that completes *amp*, *lamp*, and *clamp*, so the participant looks at those three items. The difference between Hypothesis 3 and Hypothesis 4 is that Hypothesis 3 only considers partial words that start at the beginning of the word being uttered, while Hypothesis 4 also allows partial words that start partway through the word being uttered.

Hypothesis 5: At every point, look at the most likely next word that is consistent with everything you have heard so far. For instance, for *This fish is a *clamp**: Before *clamp* has been started, the most likely word to go in that position is *carp*, so the participant looks at that image first. Then, once the *cl-* has been uttered, this word can no longer be *carp*; the most likely next word that starts with *cl-* is *clam* (as a clam is somewhat fish-like), so the participant looks at the clam. Finally, once the full *clamp* has been uttered, it rules out *clam*, so the participant switches to looking at the clamp.



(D) Made You Look! (2/3) [Solution]

D2. What trajectory will the participant's eyes follow if they view Image HH as a duck? Answer: BB,HH,II

The most likely next word, after hearing *My cousin has a pet ____*, is *cat*, so that participant starts out looking at that. Then, after the *d-* of *dolphin* has been uttered, the participant switches to the duck, since that is a more likely pet than a dolphin. And then they finally switch to the dolphin.

What trajectory will the participant's eyes follow if they view Image HH as a rabbit? Answer: B,II

Since *rabbit* does not start with *d-*, it cannot be an intermediate step between *cat* and *dolphin*. (The same holds true even if you call it a bunny or a hare.)

D3. The participant's word for a carbonated beverage: Answer: soda

The participant's trajectory went *water* -> (*carbonated beverage*) -> (*missing*) -> *sushi*. Since the carbonated beverage was an intermediate step between *water* and *sushi*, the participant must be using some word for carbonated beverage that shares its start with *sushi*. The only option that fits this description is *soda*.

D4. What should Image FF be? Answer: soup

The participant's trajectory went *water* -> *soda* -> (*missing*) -> *sushi*. The missing item must start with the sounds *soo*, in order to work as an intermediate step between *soda* and *sushi*. In addition, it has to be something that is a more likely continuation than *sushi* for *Alice was thirsty after her run, so she ordered some ____*. The most plausible thing that starts with *soo* and that can fit in that slot is *soup*.

D5. Fill in the table:

	Sentence	Trajectory
i	I bought my friend a birthday *campfire*	(a) F22,F17,F13,F08,F04
ii	(b) I went on a walk with my pet *camera*	(c) F26,F09,F18,F08
iii	(d) The queen lives in a magnificent *tennis ball*	F27,F23,F19,F14,F10
iv	My neighbor rides to work every morning in my *bow tie*	(e) F03,F06,F12,F15
v	He signed his name in ink with a dark blue *pencil*	(f) F24,F05
vi	My neighbor rides to work every morning in my *traffic cone*	(g) F03,F01,F02,F21
vii	At the construction site there was a pile of *strawberries*	(h) F07,F11,F16,F20,F25

Multiple options are possible for the sentences.

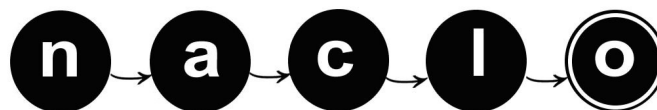
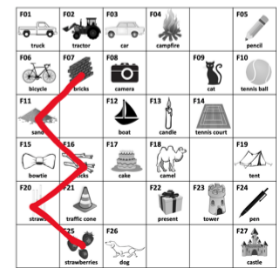
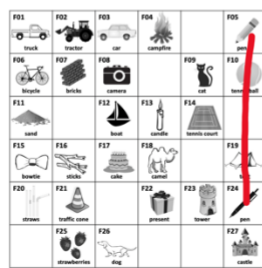
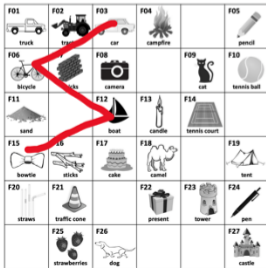
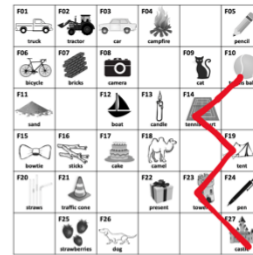
See **D6** for an explanation of the secret message, which must be used to figure out the trajectory that is needed for (ii). The trajectory F26,F09,F18,F04 is also acceptable, though for a slightly lower score, since it does not make as nice of a Y shape.



(D) Made You Look! (3/3) [Solution]

D6. What is the secret message? Answer: EYE SITE

If you trace each trajectory from Q5 on the grid, you will see that it forms a letter (see below). Taking these letters in order forms EYE SITE – an appropriate title for a problem about eye positions.



(E) The Script of Your Dreams (1/1) [Solution]

E1.

1. A.
2. H.
3. M.
4. I.
5. F.
6. B.
7. R.
8. C.
9. E.
10. G.
11. O.
12. N.
13. P.
14. Q.
15. D.
16. V.
17. L.
18. U.
19. S.
20. K.
21. J.
22. T.

E2. W

{i, u, e, o} correspond to $\mathfrak{S} + \{1, 2, 3, 4\}$ — so a corresponds to $\mathfrak{S} + 5$

5 in Mandombe:



(F) The Wiles of Reptiles (1/2) [Solution]

F1. Translate the eighth sentence into English.

Answer: The tall women carried a fearful chief.

F2 and F3. Multiple choice to fix the incorrect sentences, and translations of all six:

(a) Incorrect (verb and subject disagree; -t goes with 1p subject, -n with 3p)

Multiple choice: X (“κόχιν” → “κόχit”)

Translation: I loved a black eagle.

(b) Correct

Multiple choice: W

Translation: Tall girls deceived the selfish chief slowly.

Note: some other placements of the word “slowly”, e.g., “...slowly deceived...”, are acceptable

(c) Incorrect (suffixes are on noun, but adjective is present)

Multiple choice: Z (“zitáfisim koχsim” → “zitáf koχsimidisim”)

Translation: A man saw the pleasant oxen.

Note: some other pluralizations of “ox”, e.g., “oxes”, are acceptable

(d) Incorrect -- suffixes on adjective are ordered incorrectly

Multiple choice: Y (“s’anisubim” → “s’anubisim”)

Translation: The girl helped the black crocodile.

(e) Correct

Multiple choice: W

Translation: The small man washed them.

Note: option Y does not produce a correct sentence because of a missing high tone

(f) Correct

Multiple choice: W

Translation: They bit a fearful eagle quickly.

Note: some other placements of the word “quickly”, e.g., “...quickly bit...”, are acceptable



(F) The Wiles of Reptiles (2/2) [Solution]

Explanation of grammar:

Dime has Subject-Object-Verb word order. (In reality, there is some freedom in word order in Dime, but SOV is predominant.)

Noun phrases in Dime have Noun-Adjective order. (In reality, Adjective-Noun order is equally common, but for simplicity only N-Adj was shown.)

Verbs end in -t for 1st person subject, -n for 3rd person subject. (The -i- before the suffix marks past tense.)

Noun phrases are marked with suffixes for gender/number, definiteness (“the” vs. “a”), and case (“accusative” when used as an object, “nominative” when subject). When no adjective is present, the suffixes attach to the noun. When there is an adjective, suffixes mostly attach to the adjective (but the noun retains a plural marker where applicable):

Noun only:

	[Gender/Number]		[Definiteness]		[Case]
[Noun root] +	af (plural)	+	is (definite)	+	im (accusative)
	∅ (sing., any gender)		∅ (indefinite)		∅ (nominative)

Adjective + noun:

	[Gender/Number]		[Definiteness]		[Case]
[Adj root] +	ub (masc. sing.)	+	is (definite)	+	im (accusative)
	ind (fem. sing.)		∅ (indefinite)		∅ (nominative)
	id (plural)				

(followed by)

[Noun root] +	af (plural)	+	<i>unmarked</i>	+	<i>unmarked</i>
			<i>Singular unmarked</i>		

Note: Nouns that end in a high tone vowel in the singular retain high tone on endings, e.g., goštú → goštáf

Other:

- Animals and chiefs take masculine markings by default
- ṛóllóχ (slowly) and ṛólóχ (quickly) contrast by consonant length
- kété (they), kénim (them), and ṛaté (I) pronouns are attested



(G) The Skates of Wrath (1/1) [Solution]

Meanings of abbreviations:

DIM = diminutive (meaning roughly “little” or “young”)

LOC = locative (meaning roughly “place”)

INF = infinitive (the “to X” form of a verb X)

PL = plural

G1. Correct matches:

1. R (tell-INF)
2. B
3. N (dog-DIM)
4. G (sit-LOC)
5. Q (sit-INF)
6. E (wash-LOC)
7. I (Mexican-food-LOC)
8. C (big-wheel)
9. L (big-town)
10. J (paper-tell)
11. M
12. F (paper-LOC)
13. H
14. P (metal-tell)
15. A (metal-brain)
16. O (wheel-PL-shoe-PL)
17. K (horse-food)
18. D (horse-DIM)

G2.

tôskashâ: horse

kôkenêwa: to wash (wash-INF)

G3.

brain: wînêtepi

shoe store: mahkathêhanikâneki (shoe-PL-LOC)

also acceptable: mahkathêhikâneki (shoe-LOC)

town: otêneki

also acceptable: têneki, otênekêhi, tênekêhi, otênehêki, tênehêki (town *and* town-DIM, according to various acceptable analyses of the components)



(H) Sequitur (1/2) [Solution]

The Sequitur algorithm was introduced in 1997 by Craig Nevill-Manning and Ian Witten to “identify hierarchical structure in sequences.” Note that the version of Sequitur shown in this problem has been modified slightly from the original version (specifically, the “rule utility” property is not enforced in this problem). Byte pair encoding (BPE) was described in a 1994 paper by Philip Gage. It has been used in many of the most sophisticated natural language processing systems developed in recent years.

H1. The complete table for Sequitur is:

Step Number	Column 1	Column 2
1)	a	S = a
2)	ab	S = ab
3)	abc	S = abc
4)	(a) abcd	S = abcd
5)	abcdb	(b) S = abcdb
6)	abcdbc	S = abcdbc
	aXdX	S = aXdX X = bc
7)	(c) aXdXa	(d) S = aXdXa (e) X = bc
8)	aXdXab	(f) S = aXdXab X = bc
9)	aXdXabc	S = aXdXabc X = bc
	aXdXaX	S = aXdXaX (g) X = bc
	(h) YdXY	(i) S = YdXY X = bc
10)	(j) YdXYd	Y = aX (k) S = YdXYd
	ZXZ	X = bc Y = aX S = ZXZ X = bc Y = aX Z = Yd
S = ZXZ; X = bc; Y = aX; Z = Yd		

Brief explanation: Sequitur considers its input letter by letter, moving left-to-right. When it encounters a two-letter-sequence that it has seen before, it replaces all instances with a new letter (“nonterminal”, e.g., X), and makes any other possible replacements of repeating sequences with nonterminals. Column 1 records the portion of the input under consideration at each step (including the results of nonterminal replacements) and Column 2 keeps track of the nonterminal-to-letter correspondences in replacements.



(H) Sequitur (2/2) [Solution]

H2. The complete table for BPE is:

Step Number	Column 1	Column 2
1)	abcdbcabcd	S = abcdbcabcd
2)	(a) aXdXaXd	(b) S = aXdXaXd X = bc
3)	YdXYd	S = YdXYd X = bc (c) Y = aX
4)	ZXZ	S = ZXZ X = bc Y = aX Z = Yd
S = ZXZ; X = bc; Y = aX; Z = Yd		

Brief explanation: BPE considers the entire input at once, and determines the most-commonly-occurring two-letter sequence across the whole input (breaking ties in favour of the leftmost sequence). This sequence is replaced with a nonterminal, and the process repeats, continuing until no replacements are possible. Columns 1 and 2 serve purposes analogous to those in Sequitur.

H3.

- (a) Different *Sequitur* output: S = XcXdYY; X = ab; Y = bc
 BPE output: S = aXabdXX; X = bc
- (b) Same *Sequitur* and *BPE* output: S = abbXddX; X = ca
- (c) Different *Sequitur* output: S = VV; X = cb; Y = ba; Z = YX; W = ZX; V = Wa
 BPE output: S = VV; X = ba; Y = Xc; Z = Yb; W = Zc; V = WX

Note: For symbols after Z, we use W, then V, and so on (going backwards), but any choices of new, distinct symbols are consistent with the data shown.

- (d) Same *Sequitur* and *BPE* output: S = ZZXaX; X = cc; Y = Xd; Z = Yb
- (e) Different *Sequitur* output: S = ZZWWW; X = cc; Y = Xd; Z = Yb; W = Xa
 BPE output: S = WWYYY; X = cc; Y = Xa; Z = Xd; W = Zb

Note: since these two outputs have the same "structure", that is, they are differentiated only by the specific nonterminal symbols used, the answer "Same" was given partial credit. Also, it is consistent with the data to suppose that both algorithms will make one more compression, on the final three characters (e.g., Sequitur might end with S = WWVY, with V = YY). This will not affect the answer here.



(I) Non Sequitur (1/1) [Solution]

11. Fill in the blanks:

- (a) shortest
- (b) L1 or L2
- (c) longest
- (d) L1
- (e) longest
- (f) L1 or L2
- (g) longest
- (h) L2
- (i) longest
- (j) L1 or L2
- (k) longest
- (l) L1 or L2

12. Definitions (multiple answers listed under a given strategy are all acceptable):

- (a) reunion
1. the opposite of a charged particle again
 2. a workers' organization again
 3. a workers' organization again
the opposite of a charged particle again
 4. a workers' organization again

- (b) unrest
1. the opposite of a street again
 2. the opposite of stay still
 3. the opposite of stay still
the opposite of a street again
 4. the opposite of stay still

- (c) presto
1. a street again (??)
 2. stay still (??)
 3. not yet a street (?)
 4. stay still (??)

13. Other word: recent

Definition: a penny again

