The United Kingdom Linguistics Olympiad

Expanded Problems
Japaneasy

(1) What language(s) does this problem involve?
Japanese

What is the aim of this problem?
To work out how the Japanese hiragana writing system works.

(2) Background Information

Japanese is spoken by approximately 121 million people, in Japan. It is the largest language in the Japonic language family, but has a number of different dialects.

Japanese is written using a variety of scripts including Chinese characters. For the purposes of learning to read, however, and in the case of unusual words, it is quite common to use a phonetic system, *hiragana*.

(3) The problem

On the left below are eight examples of simple Japanese words written in *hiragana*, with their pronunciation in a random order on the right. English meanings are also given but for information only – they play no part in solving this problem.

1. あか	A. kita ‘north’
2. さと	B. kao ‘face’
3. でんき	C. denki ‘electricity’
4. あさ	D. tsudo ‘every time’
5. かつどん	E. asa ‘morning’
6. かお	F. katsudon ‘pork and rice’
7. きた	G. aka ‘red’
8. つど	H. sato ‘village’

2.1. (4 marks) Show how the words on the left correspond to the pronunciations on the right:

Japanese	1	2	3	4	5	6	7	8
English								

2.2. (3 marks) How would you pronounce the following Japanese words?

(a) おでき ‘eruption’

(b) だん ‘group’

2.3. (3 marks) How would you write the following words in *hiragana*?

(a) satsu ‘banknote’

(b) kanten ‘point of view’

4) Solutions and mark-scheme

2.1	**Japanese**	1	2	3	4	5	6	7	8	½ each
	English	G	H	C	E	F	B	A	D	
2.2	**a**	odeki								1
	b	dan								2
2.3	**a**	さつ ½ for each character correct.								1
	b	かんてん ½ for each character correct.								2
	Total									10

5) Commentary

Exploring the system

The longest Hiragana form presumably corresponds to the longest pronunciation; so let's assume for now that Hiragana 5, with four characters, has pronunciation F:

かつどん = katsudon

Many UKLO questions ask you to work out how a language works, from just a small set of words or sentences. In linguistics we call the set of words or sentences the **data**.

Now, how can you use **four** characters to represent a pronunciation written with **eight** letters?

Guess 1: each character shows a syllable: a combination of a consonant followed by a vowel. So the first Hiragana character means 'ka' and the second 'tsu'. That leaves a single syllable, 'don', with two characters: かつどん = katsudon

This 'guess' is obvious if you already know that some writing systems work in this way. The more UKLO questions you do, and the more languages you come into contact with, the more background knowledge you will gain to help you with new questions.

Guess 2: all the other syllables in the word are 'open' (i.e. there's no final consonant), so maybe the final consonant 'n' gets a separate character. If so, the third character is 'do' and the fourth 'n': かつどん = katsudon

In linguistics we call a 'guess' a hypothesis. You *test* the hypothesis by stating it clearly then looking at the data to see if it is right.

The guess about 'n' is immediately confirmed when you look for other words containing [n], and you find just one pronunciation (C. denki) and one Hiragana form with the 'n' character in just the right place:

でんき = denki

If your hypotheses are right, they'll guide you to a sensible solution. They may be wrong though, so keep an open mind until you've checked all the data.

Now you've cracked the characters in [katsudon] the other characters should be easy to pin down. From just [katsudon] and [denki] you can get a long way:

あか = ??ka
さと =
でんき = denki
あさ =
かつどん = katsudon
かお = ka??
きた = ki??
つど = tsudo

What do you know so far?

Grammar: in Hiragana some symbols represent whole syllables and some represent single sounds.

Vocabulary - symbols we know so far:

か = ka　　つ = tsu

ど = do　　で = de

ん = --n　　き = ki

From those you can use the list of possible pronunciations to identify more characters

e.g. <あ> = [a].

あか = aka
さと = must be sato
でんき = denki
あさ = a?? must be asa
かつどん = katsudon
かお = kao
きた = kita
つど = tsudo

Then you can deduce the rest.

Vocabulary - more symbols we now know:

あ = a　　さ = sa

と = to　　お = o　　た = ta

Grammar: Somewhere around this point you might notice that one of the symbols appears in two forms, with/without extra marks on it: と to ~ ど do

At this stage you don't know whether this detail is important or not, so maybe make a note of it, just in case. Linguistic analysis = working out patterns in language: it is a balance between noticing details and seeing the overall pattern.

Answering the questions

Most of the questions can be answered on the basis of the available data. BUT two can't: 2.2b and 2.3b.

Both 2.2b and 2.3b raise the same difficulty: how to go beyond the basic data to see if you understand the *principles* behind the correspondences between characters and pronunciations.

2.2b includes a Hiragana character that's not in the basic data, but it's built out of the [ta] symbol plus a little equals sign (=): た ta ~ だ ??. If you didn't already spot it earlier, you should now notice you get the same pattern between [to] and [do]: と to ~ ど do. You also get one of these symbols on [de]: で.

Now you have to look for a pattern, to complete the analogy: the plain symbols all have a [t] sound, and the ones marked with the (=) sign all have a [d]:

と to ~ ど do　　た ta ~ だ ?? must be た ta ~ だ da　　?? de ~ で de must be て te ~ で de

Many UKLO questions take you up onto a higher level of difficulty by going beyond the basic data in this way. What you are always looking for is some kind of underlying pattern or rationale in the system: step 1 takes you from basic data to part of the system, then step 2 takes you, via general patterns, to other parts of the system.

6) Taking it further

Going further 1: languages are rule-governed

The linguistic concept that is the key to this question (and most other UKLO questions) is the following: languages are to a large extent systematic and 'rule-governed'. That is, you can work out some general rules for how to 'do' things in language, and once you have done so, you can *generalise* from the data you have (the words/sentences you already know) and apply the rules to new data (new words or new sentences). We at UKLO use this concept to create fiendish test questions, but it is also the basic premise of all linguistic analysis. More importantly, as human language users and learners, we all work on these principles all the time. Here is a sequence of tasks you could use to explore these ideas further and build language awareness (e.g. in a SPaG oriented lesson).

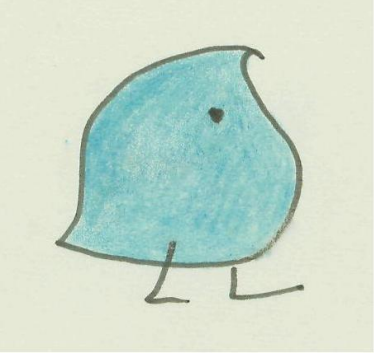
Q1: How do we make plurals in English? A: We add an 's' to the words, to make it a plural.
Q2: Does this 'rule' work for all words in English? A: Most, but not all e.g. child~children.

- Can you think of other exceptions to the rule? (mouse/mice etc etc)

Q3: to make a plural in writing you add a letter 's', but do you always *pronounce* the letter as an [s] sound?

- *Hint*: say these words slowly and stretch out the last sound: <cats> ~ <dogs>. You should hear something like: [katsssssssss] ~ [dogzzzzzz]
- Have you ever noticed this before? [Students who have had formal instruction in English as a second or foreign language may have been taught this rule].

The Wug Test was a classic experiment run in the 1950s with primary school age children in the USA, to see whether languages are really rule-governed, and at what age children 'learn' the rules. The children were shown a picture of a strange new creature and were told that it was a 'wug'; then they saw a picture of two of the creatures and were asked to fill in the blank in this sentence: "This is a wug. Now there are two of them. There are two ____." The experimenter listened to the answer and wrote down how each child pronounced the word.



Q4: What do you think the answer is? Should it be [wugssssss] or [wugzzzzzz]? A: [wugz]
Q5: Do you think the children knew the answer? A: They had never been taught how to say the plural of [wug], but they got the answer right, most of the time (i.e. they pronounced the word [wugz], following the 'rules' of English pronunciation).

You can download the original pictures here: http://childes.talkbank.org/topics/wugs/

- the file wugpicts.zip contains the pictures
- the file wug.pdf is the scientific article with the results of the experiment

Going further 2: language diversity in writing systems

Another linguistic concept which we exploit in this puzzle is the fact that languages are diverse. All human languages solve the shared problem of getting the ideas out of one person's head and into the head of another person, via a code which is external to both heads. In speech the code comprises sequences of sounds; in a sign language the code comprises gestures of the hands or body in time and space; in written language, the code comprises graphical symbols, which correspond to the sounds and words of a spoken language.

Different languages have solved the problem of how to represent the words/sounds of their spoken language in written symbols in different ways. We are very familiar with the fact that English spelling is not fully phonetic – that is, the way a word is written is not always a reliable guide to it is pronounced. Japanese hiragana is an example of a syllable-based writing system. Sequences of sounds are not always represented by sequences of symbols; the sounds in a single syllable can be represented as one unit, as we saw: と to ~ た ta ~ て te or か.

Q: What other writing systems do you know about? e.g. hierogplyphics

- Which of these writing systems are 'alphabets' like English (with symbols for sequences of consonants and vowels)?
- Which of them are syllable-based writing systems, like Japanese? (e.g. Chinese)
- Can you find an example of an 'abjad' writing system? This is one which mostly only has symbols for the consonants in the word. (e.g. Hebrew/Arabic/Farsi/Urdu)
- *Advanced*: Can you think of examples of words written in English without vowels, e.g. in text messages? Are all vowels left out, or only some? Is there a pattern to which vowels get left out. [Hypothesis: this is related to how often the word is used (frequency), and how long the word is (word length).]

Going further 3: language diversity in permitted sound sequences

Languages also differ is what sequences of sounds can go together to form a word (the technical term for this is 'phonotactics' – which sounds can 'touch').

Japanese is very different from English in this respect. English permits 'clusters' of consonants, as the beginning and end of words, e.g. "splits". Japanese does not allow consonant clusters, and only allows a small set of nasal sounds to occur at the end of syllable, such as [m] and [n]; most syllables in Japanese are 'open' (that is comprising a consonant + vowel only).
Task: are these patterns true in the words we have given you in the UKLO question? {A: they should be!).

These patterns are reflected in the writing system. A closed syllable (e.g. ending in [n]) is an unusual thing in Japanese, and it therefore gets a symbol of its own in the writing system.

When English words are borrowed into Japanese, the word generally gets changed to fit the rules of Japanese: e.g. 'bus' = [basu]; 'grey'= [gure]; 'Beckham' = [bekuharu].

When we borrow Japanese words into English we don't change them as much, as our rules about what sounds can occur are not so strict as those of Japanese.
Task: can you think of some Japanese words we use in English? [A: sushi, wasabi, kimono, tsunami].