(C) Orthography Design

Orthography design is the process of developing an alphabet and spelling rules for a language. A good orthography has several features:

> Given a spoken word, there's no question of how to spell it.
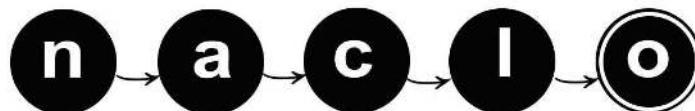> Given a written word, there's no question of how to pronounce it.
> In the modern world, it's increasingly important that it be reasonably easy to type!

Quechua is spoken today by millions of people in Peru, Ecuador, and Bolivia, the descendents of the citizens of the Incan Empire. Quechua speakers are rapidly joining the Information Age, and both Google and Microsoft Windows now come in Quechua!

Like in English, there are more sounds in Quechua than there are letters on a keyboard, but there are ways around that. For example, we can assign one letter to multiple sounds so long as a reader can always predict, from its position in the word or from other letters in the word, which sound is meant. So if the sound [b] only ever occurs right after [m], and [p] never occurs right after [m], we can just write "p" for both, since you'll be able to predict from the previous letter whether "p" means [b] or [p].

This "phonemic principle" is the central principle of most orthographies, not just because it reduces letters but also because our minds categorize sounds in the same way.

Here are 18 words in Cuzco Quechua, as they are pronounced but not necessarily as they are written. [q] and [χ] represent special sounds that don't occur in English.
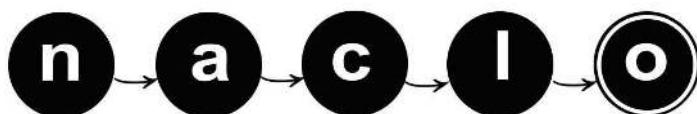
(C) Orthography Design

awtu	car	**qasi**	free	**seqay**	to climb
kanka	roasted	**qatoχ**	merchant	**sikasika**	caterpil-lar
karu	far	**qatuy**	to barter	**sipeχ**	mur-derer
kiru	teeth	**qatisaχ**	I will fol-low	**sipiy**	to kill
kisa	nettle	**qelqaχ**	writer	**soχtaral**	sixty cents
kisu	cheese	**qelqay**	to write	**sunka**	beard
kunka	neck	**qolqe**	silver	**toχra**	ball of ash
kusa	great	**qosa**	husband	**uyariy**	to listen
layqa	witch	**qosqo**	Cuzco	**uywaχ**	care-taker
oqe	spotted	**saqey**	to abandon	**waleχ**	a lot
qasa	frost	**saχsa**	striped	**weqaw**	waist

Notes:

• It is quite expected that few if any contestants are going to get all 20 points. There are going to be entirely correct and well-explained answers that don't get quite as many points as another entirely correct and well-explained answer because the latter was more thorough. (In the first version of this rubric, we found that the minimal "correct score" was about 12, give or take.)

• Some solvers will have completely misunderstood what they were supposed to do. This is too bad, but they don't get any points for well-meaning but bizarre answers! This is a contest, rather than a homework assignment, and for some of the puzzles the puzzle *is* to figure out what's being asked.

n a c l o

(C) Orthography Design

- Half-points may be awarded.
- It is *not* necessary for a complete solution that the solver to chooses <u> and <i> to be basic rather than <o> and <e>. From a phonemic point of view, the label of a sound is arbitrary – these could be Smiley Face and Labialized Smiley Face for all we care – and from an orthographic point of view, it doesn't matter, they're all just symbols.

I. Show that we don't need separate letters for [q] and [χ]. (3pts)

- 1a. **1pt**. for noticing that they never occur in the same environments
- 1b. **1pt**. for correctly specifying what these environments are.
- 1c. **1pt**. for clearly explaining why this means they can be the same letter. (This explanation doesn't have to be *long*, just clear.)

II. Show that we can't represent [a] and [i] by the same letter. (3pts)

- 2a. **1pt**. for noticing that they do occur in the same environments
- 2b. **1pt**. for finding a minimal pair, like "karu ~ kiru" or "qasa ~ qasi". (If they have this but not 2a., give them the point for 1a anyway, since this subsumes that.)
- 2c. **1pt**. for clearly explaining why this means they have to have different letters.

III. Show that we can't represent [a] and [e] by the same letter. (3pts)

- 3a. **1pt**. for noticing that they do occur in the same environments.
- 3b. **1pt**. for noticing the pair "saqey" ~ "seqay". (If they have this but not 3a, again give them that point anyway.)
3c. **1pt**. for clearly explaining why this means they have to have different letters.

IV. Most modern Quechua orthographies get by with only three of the five vowels [a], [e], [i], [o], and [u]. Show how this is possible. (11pts)

(C) Orthography Design

First, they should establish which sounds *can't* be merged into a single letter:

- 4a. **1pt**. for finding a pair kisa ~ kisu or kanka ~ kunka.
- 4b. **1pt**. for finding the pair qasa ~ qosa.
- 4c. **1pt**. for finding the pair kisa ~ kusa.
4d. **1pt**. for recognizing the relevance of the pairs in II and III to this question.

Second, three points for deducing which sounds can be merged:

- **1pt**. for figuring out that either [e]~[i] and [o]~[u], or [e]~[u] and [o]~[i]. (They don't have to get both for this point.)
- **1pt**. for figuring out that both of these are possible.
- **1pt**. for clearly explaining how this follows from the facts above and in parts II and III – that given the minimal pairs, these two are the only solutions that don't cause two different words to be spelled the same.

Four points are available for:

•Up to **2** points for determining the conditioning environment for the difference: [e] and [o] when next to [q] and [χ], [i] and [u] elsewhere. (1 point each for the completeness of the description and the clarity of the explanation.)

•Up to **2** points for determining, based on the alternations *qelqay ~ qelqaχ*, *qatuy ~ qatoχ*, and *sipiy ~ sipeχ*, that [o]~[u] and [e]~[i] is the better or more likely of the possible solutions. (1 point for noticing the pattern and 1 point for correctly deducing the right phonemicization.)

n a c l o