

In-Class Activity

Language Identification

Wednesday, January 30, 2008

Using the trigram frequency distribution for Czech and Polish on the back of this handout, you want to identify the language of the following phrase:

A t o c i l e c i !

1. In the following table, list all the trigrams in the phrase above:

Trigram	Czech	Polish

Trigram	Czech	Polish

2. Using the table on the back, fill out the rest of the table and figure out which language this method predicts for the phrase above.
3. Looking at the table on the back again, find one character that is only used in Czech and one that is only used in Polish.