Solution to the "Dorabella" Cipher

Tim S Roberts, CQUniversity, Bundaberg, Queensland 4670, Australia

This is a simple one-page justification of the solution given at http://unsolvedproblems.org/S12x.pdf

Suppose the first 50 or so characters are translated into plain text, but this is purely by chance, or some connivance (the odds against such a possibility are huge, but never mind). What are the chances, for example, of a 5-letter name, such as ***Luigi***, appearing amongst the remaining characters?

Suppose there are 1000 recognisable 5-letter names. In the remaining 37 characters of the message, there are some 33 places in which such a name could appear. Each ciphertext symbol can have 24 possible plaintext interpretations. The chance of one of the 1000 names appearing by chance, therefore, is 1000*33 / 24^5, which is 33000 in 7962624, or 1 in 241. Odds of 240 to 1 against.

Now, this is large odds against, but is certainly possible. However, including other meaningful words appearing by chance amongst the remaining characters increases these odds by many orders of magnitude. Having the name ***Luigi*** (Italian) and also the Italian word ***liuto***? Well, let's work that out. Suppose there are 1000 such Italian words. Then there are 32 remaining characters for this word to appear, and we have odds for this single Italian word (or 1000 others like it) to occur of 32 * 1000 / 24*24*24*24*24 = 1 in 248. For both an Italian name and an Italian word to occur we therefore have odds of 240 * 248 = 59,520 to 1 by random chance.

OK, what about a third Italian word, such as ***studo***, occurring within the same text? Well, about 1 in 240, as before. So for all three such words (***Luigi, liuto, studo***) to occur, the odds against are approximately 59520*240 = 14,284,800 to 1.

Now, suppose the message has a single letter deliberately omitted. What are the odds of the apostrophe occurring at exactly this point? Roughly 1 in 87.

What are the odds against all of the above occurring by chance? 14,284,800*87 = 12,427,776,000 to 1 against.

And we now have to consider other words (such as "***luv'ngly tuned***"), and any meaningful key, arising from the above, by pure chance. This is difficult to estimate, but at a minimum, but would be many millions to 1 against. Whatever estimate is chosen, this must be multiplied by the estimate above.

This is far higher than any DNA evidence required to convict someone of first-degree murder, beyond reasonable doubt. We therefore conclude that the solution is correct, beyond reasonable doubt.

Tim S Roberts

April 2011