

A mathematician's view of the use of mathematics in George's "Long o-type vowels in Cornish"

Dr Mícheál Ó Searcóid
Department of Mathematics
University College, Dublin
1997

Introduction

I have been asked to make comments on the use of mathematics in Dr George's article "Long o-type vowels in Cornish". I should say at the outset that I have no knowledge of the Cornish language and am not versed in linguistics or in the analysis of texts. Nonetheless, I have read the article with interest and consider myself obliged to make some general comments before looking specifically at the mathematics.

The problem, as I understand it, is as follows: In Old Cornish there were two distinct phonemes, represented here, as they are in Dr George's article, by /ɔ/ and /o:/, the development of which is disputed. Dr Williams contends that they fell together and Dr George that they did not.

I understand that the existence in Old Cornish of these two distinct phonemes is not in dispute and that there is complete agreement on which vowels of Middle and Late Cornish have their origins in each of these two separate phonemes, irrespective of which grapheme is used to indicate each vowel. This understanding must be stated, as it is relevant to what I shall later say about the mathematical analysis. The correctness of my analysis depends on the correctness of this understanding.

Whether or not the symbols /ɔ:/ and /o:/ accurately represent the phonemes as they existed in Old Cornish is irrelevant to the mathematical problem and I use them here without prejudice simply for convenience.

Two sets of graphemes, called here, as in Dr George's article, X type and Y type spellings, are used in the texts to indicate variously the two phonemes. I must assume, for lack of any specific knowledge, that the division of graphemes into these two sets is appropriate. So I will follow Dr George and work with the two sets of graphemes rather than with the individual graphemes which appear in the texts.

The use of mathematics in text analysis

Particularly since the advent of modern computer software, it has become possible to make detailed computational analysis of texts in order to raise or to answer questions relating to the texts or their language. Such analysis is certainly valid.

There are some tasks which hardly require computers. For example, a few pages taken at random from modern English texts will display fairly accurately the distribution of the use of the 26 letters in most random texts. In this case a small and easily analysed sample is sufficient to provide accurate results about character frequency in the written language. The number of texts available for analysis by Dr George is small, but I do not think that this in itself is sufficient reason to discount any valid calculation made on the basis of those texts—whether that calculation be linguistic or mathematical. Let the enquiring mind squeeze whatever knowledge it can from all the available sources.

Dr George's reason for believing that the two phonemes persisted into middle and late Cornish seems to be based mostly on counting; he has made lesser play of linguistic

considerations and none of social or psychological ones. It is not clear to me that counting can give anything more than a rough indication. There are too many other things to be considered. Even a non-linguist must raise questions such as the following:

1. Was the writer the author or merely a scribe?
2. At what period was the text originally written and, if it is a copy, at what date was it copied?
3. What part of Cornwall did the writer come from?
4. What formal training and knowledge of literature did the writer have?
5. Was the writer a native speaker of Cornish?
6. What influence might the English language have had on the text?
7. What comparisons can be made between the development of the Cornish language and other Celtic languages?
8. What comparisons can be made between the development of the Cornish language and other languages which are either dying or coming heavily under the influence of a foreign language?

These questions I throw out at random. I could list many more; the point is made, however, that orthography does not necessarily represent the state of the living language spoken by the writer. It is hardly necessary to go beyond modern English to illustrate this.

It is questions such as these which lead me to believe that counting, however thoroughly executed, is never going to be any better than a very crude tool to deal with the problem at hand. The problem belongs more in a linguistic, social and historical context than in a mathematical one.

Dr George's figures

What, then, can we draw from Dr George's figures? There are two assertions which would be difficult to dispute on the basis of these figures: Firstly, that X type spellings were much more common in written Cornish than were Y type spellings. Secondly, that instances of use of the /o:/ phoneme in the language were rare compared to instances of the /ɔ:/ phoneme. The question of whether or not the figures suggest a falling together of the phonemes is more difficult, and I am not sure that the difficulty lies in the paucity of the material.

In addressing this question, let us make two other observations clearly indicated by the figures.

The first is that a consensus exists in the texts that the phoneme /ɔ:/ carries an X type spelling. The only text which deviates substantially from this consensus is *Beunans Meriasek*, but even in that the preference is for the X type spelling.

The second is that there is no consensus about the precise orthography to be used to indicate the phoneme /o:/. In the list given on page 11, this phoneme is spelt with an X type spelling in 369 instances and with a Y type spelling in 275 instances. Moreover, there is no evident pattern of development from one spelling type to the other over the period of time in question.

I would suggest that the figures for /o:/, particularly when compared to those for /ɔ:/, might lead one to consider the possibility that the /o:/ phoneme itself did not have an unchallenged position in the language throughout any of this period. Perhaps, indeed, they indicate that /o:/ as a distinct phoneme was disappearing or had disappeared, thus supporting rather than refuting Dr Williams' conjecture.

The mathematical computation

The expression given by Dr George on page 9 calculates the probability that, given a population of P members in which D members exhibit a particular property, a set of p members randomly chosen from the population will contain exactly d members which exhibit that property. In his example, Dr George takes P to be the total number of words analysed, which is 343; he takes D to be the total number of those words with a type Y spelling, namely 13; and he takes p to be the total number of words which exhibit the /o:/ phoneme, namely 49. He must then perform the calculations for $d=10$, $d=11$, $d=12$ and $d=13$ and add these together to calculate the probability that a set of 49 members randomly chosen from the population will contain at least 10 with Y type spellings.

So far this is correct. The major flaw in using such a calculation to throw light on the question raised by Dr George is that the particular set of 49 members is not randomly chosen. It is precisely the set of words which exhibit the /o:/ phoneme. It was never the contention of Dr Williams, nor could it have been, that the etymology of the words disappeared when the two phonemes fell together. If we suspend our knowledge of the origins of the various words and regard the total population as being simply a set of 343 words of which 330 exhibit an X type spelling whilst the other 13 exhibit a Y type spelling, then it is highly unlikely—as Dr George asserts—that a random set of 49 of the words will contain at least 10 of the 13 Y type spellings. When a particular set of 49 does indeed contain 10 of them, it is reasonably safe to assume that it was not randomly chosen. And indeed, in this case, it was not, for there is an undisputed historical difference between these 49 and the rest.

I must, therefore, on my understanding of the problem stated at the beginning of this note, conclude that the problem has been incorrectly transcribed into mathematics and that the calculations have no bearing whatsoever on the solution of the problem.

One last question

I have said that computational analysis can be used not only to help in solving problems, but also to raise questions. It has been noted by Dr George that the figures for *Beunans Meriasek* are out of step with those for the other documents. If there is one question clearly raised by these figures it is this: Why did Radulphus Ton deviate so greatly from the clear consensus shown in all the other texts that /l:/ takes an X type spelling? I do not understand Dr George's answer, which seems to me dismissive. Perhaps this is because I do not know the text and am not versed in these matters generally. But I cannot help observing that idiosyncratic people emerge in every age; perhaps Radulphus Ton was a soul mate of his fellow Celt, T. E. Lawrence, who, when asked by the publisher of *Revolt in the Desert* to help in cleaning up the inconsistent spellings of proper names in the text, wrote "I spell my names anyhow, to show what rot the systems are."