Continuous modelling of verse lengths in Welsh and Gaelic metrical psalmody

Andrew Wilson
Lancaster University

Abstract
Twenty 17th century metrical psalms (ten in Welsh and ten in Scottish Gaelic) were used to investigate the suitability of the Zipf-Alekseev function for modelling verse lengths in these Celtic languages. Good fits were obtained for both languages; however, the parameter *c* was always negligible, hence future work might consider using the simple power function, if this proves to be the case more generally. Future work needs to look at a broader sample of authors, dates, genres, etc. in order to arrive at a full synergetic model for verse lengths in Welsh and Scottish Gaelic.

Key words: Welsh; Scottish Gaelic; verse lengths; Zipf-Alekseev function; power function.

1. Introduction

Poetry can come in many forms but the commonest involves sets of verses with a fixed metrical pattern of long and short, or stressed and unstressed, syllables. Such poetic metres allow very limited scope for variation in the lengths of lines when they are measured in syllables and deviations from the metre occur only rarely. The only exceptions to this limit on variability are those metres where a certain number of syllables of a particular type may be substituted for a different number of another type - e.g., in the classical Latin hexameter and pentameter, where a choice of either two short syllables or one long syllable is allowable in many positions. However, if we measure the lengths of verses in words, rather than in syllables, then there is more scope for variation. Nevertheless, the constraints of the metrical form, and the word-length structure of a language's lexicon, will still have some effect on the number of words in a verse. What we have here, then, is a potential stochastic regularity that enters into a larger synergetic linguistic system (Köhler, 1987).

Verse length, in words, has not yet been extensively researched for any language. Best (2012a) cites only two previous works by Muller (1972) and Grotjahn (1979). However, in a recent series of studies, Best (2012a, 2012b, 2013) has begun to explore verse lengths in German and Old Icelandic using discrete probability distributions. He has found that, so far, the 1-displaced Binomial distribution constitutes a good model for the verse lengths in both languages. However, a problem with discrete modelling, as noted in the context of word-length research by Mačutek and Altmann (2007), is that many different distributions may be required to cover the world's languages, and different distributions may even be required within a single language to account for variations in date, text-type, etc. Whilst most of these distributions may be derivable as special cases from a more general law, their proliferation complicates the picture, especially when one seeks to interpret parameter variation from a functional perspective. But discrete modelling is not the only form of modelling that can be applied to language and texts. Mačutek and Altmann (2007) have noted that discrete and continuous modelling are merely two sides of the same coin. Since all modelling involves an idealized abstraction of entities and properties from concrete reality, it is also possible to consider essentially discrete phenomena from a continuous perspective - which is the case, for example, in the modelling of rank-frequency relations using Zipf's power law. Using a continuous function in place of a discrete distribution opens up the possibility of arriving at a unique model of a phenomenon, where only the parameters of the function are free to vary from text to text, regardless of language or other typological considerations. Having once established the general appropriacy of the model, the nature and grounds of any variation in the parameter values can then be studied more systematically across different data sets.

Popescu, Best & Altmann (2014) thus began to experiment with a single continuous function - the Zipf-Alekseev function – to study the distribution of lengths for different units in texts. This function derives from the differential equation:

$$dy/y = (A + B \ln x / Dx)\, dx$$

where x is the length class and y is the frequency of the length class. A is the language, text-type, or style constant; B is the force of the speaker or writer; and D is the equilibrating force of the community (Koch, 2014). Solving and re-paramterizing this equation gives the Zipf-Alekseev formula:

$$y = ax^{b+c \ln x}$$

The aim of this paper is to provide a preliminary test of whether this model can also account for the distribution of verse lengths in Celtic poetry, specifically within the genre of metrical psalmody in Welsh and Scottish Gaelic.

2. Data and method

This study is the first to consider Celtic verse lengths from the theoretical perspective of Popescu, Best & Altmann (2014), hence the focus is on establishing the appropriacy of the model, rather than on investigating patterns of parameter variation across authors, genres, dates, etc. Thus, in order to minimize predictable variation within the data samples for this pilot study, the scope of the experiment was restricted to 17th century metrical psalmody, a poetic genre that exists for both Welsh (a P-Celtic or Brythonic language) and Scottish Gaelic (a Q-Celtic or Goidelic language). Within the individual language samples, the metrical psalms considered here allowed date, metre, genre, and - at least for Welsh – authorship to be held constant, so that any language-internal parameter variation or failures in model fit must be attributable to other, less immediately obvious factors. Also, since the data are free translations of the biblical book of Psalms, some of which are quite lengthy, it was possible to obtain a reasonably sized corpus of sufficiently long texts that adhere to these criteria, something that is not always possible with poetic genres that favour short texts. If we were to rely on other texts at this stage in the research, we could encounter difficulty in finding sets of sufficiently long texts that do not also exhibit substantial internal variation in authorship, metre, date, etc.

The Welsh metrical psalms used here are all by Edmund (in Welsh: Edmwnd) Prys, a noted Welsh academic, clergyman, and poet (Williams 2004, Morgan 2011). They were published in 1621 as an appendix to the Welsh version of the Book of Common Prayer, and all use a metre of alternating 8 and 7 syllable verses.

The Scottish Gaelic texts are of less certain authorship. Their textual history is somewhat complex, but the basic facts are these. According to MacTavish (1934), the first 50 psalms were originally translated into verse by a group of three translators and published in 1659. The rest of the psalms, together with revisions of the first 50, were completed by a different team of five translators. The authors of the individual psalm translations are, unfortunately, not identified in the text. The text, as published, dates from 1694. All of the psalms use a metre of alternating 8 and 6 syllable verses.

The texts chosen for both languages were psalms 18, 22, 37, 68, 78, 89, 104, 106, 107, and 119. These were selected as being the ten longest psalms in the Welsh translation (which was analysed first).

For the purposes of this experiment, a word was considered to be any one or more alphabetic characters with a space, punctuation mark, or line break on either side. Apart from removing blank lines, title lines, verse numbering, etc., no attempt was made to modify the orthography of the texts, and the frequencies of line lengths (in words) were counted for each text using a reliable program written in Python. The Zipf-Alekseev function was then fitted to the frequency data from each text using TableCurves. A fit was considered acceptable if it achieved the usual benchmark R^2 of 0.8.

3. Results

The results for fitting the Zipf-Alekseev function to the Welsh data are shown in Table 1, and the results for the Scottish Gaelic data are shown in Table 2.

Table 1
Welsh data

Welsh Psalm 18			Welsh Psalm 78		
Words	Verses	Z-A	Words	Verses	Z-A

2	1	1.12	3	11	9.09
3	5	9.30	4	50	51.56
4	42	38.36	5	74	71.96
5	49	52.08	6	45	47.39
6	40	38.90	7	22	20.79
7	22	20.97	8	9	7.57
8	9	9.60	9	1	2.87
a = 23.8415, b = -7.4907, c = 2.95399985E-007, R^2 = 0.9811			a = 29.9705, b = -9.4974 c = 3.8600554E-009, R^2 = 0.9946		

Welsh Psalm 106			Welsh Psalm 22		
Words	Verses	Z-A	Words	Verses	Z-A
3	14	11.39	3	3	4.41
4	37	36.16	4	29	28.11
5	37	44.69	5	43	43.87
6	47	33.48	6	31	30.59
7	13	19.14	7	15	13.82
8	4	9.56	8	3	5.22
a = 20.1539, b = -6.4036 c = 5.71773026E-006, R^2 = 0.7848			a = 32.2565, b = -10.0819 c = 2.67570099E-010, R^2 = 0.9925		

Welsh Psalm 89			Welsh Psalm 107		
Words	Verses	Z-A	Words	Verses	Z-A
2	1	1.01	3	14	8.28
3	6	5.09	4	25	29.92
4	35	35.33	5	43	39.06
5	59	58.76	6	27	29.43
6	43	43.49	7	20	16.46
7	22	20.61	8	3	7.98
8	6	7.86			
a = 32.0086, b = -9.9062 c = 3.40804279E-010, R^2 = 0.9977			a = 22.1484, b = -6.9826 c = 9.00508214E-007, R^2 = 0.8720		

Welsh Psalm 37			Welsh Psalm 104		
Words	Verses	Z-A	Words	Verses	Z-A
3	2	5.08	3	8	5.10
4	32	28.31	4	27	25.01
5	38	43.61	5	33	39.87
6	39	32.44	6	43	33.40
7	14	16.23	7	16	19.54
8	3	6.74	8	5	9.48

a = 29.0385 b = -9.0277 c = 3.07743607E-009 R^2 = 0.9206	a = 25.4990 b = -7.7907 c = 3.39436242E-008 R^2 = 0.8336

Welsh Psalm 119			Welsh Psalm 68		
Words	Verses	Z-A	Words	Verses	Z-A
3	16	11.81	3	6	6.88
4	88	87.50	4	37	35.92
5	138	141.69	5	45	46.79
6	108	101.83	6	31	28.80
7	45	46.57	7	12	12.00
8	9	16.67	8	1	4.39
a = 31.7840202 b = -9.88209848 c = 1.11953174E-009 R^2 = 0.9904			a 30.3977335 b -9.74183401 c 2.35895246E-009 R^2 = 0.9869		

Table 2
Gaelic data

Gaelic Psalm 18			Gaelic Psalm 78		
Words	Verses	Z-A	Words	Verses	Z-A
3	5	1.50	2	1	1.00
4	19	23.15	3	2	2.66
5	74	69.22	4	32	31.72
6	46	53.45	5	79	78.57
7	28	19.33	6	68	70.39
8	4	4.99	7	41	35.34
			8	8	12.89
			9	1	4.30
a = 52.931569 b = -15.9864633 c = 6.61118031E-018 R^2 = 0.9491			a = 39.3325563 b = -11.7439952 c = 4.06266085E-013 R^2 = 0.9894		

Gaelic Psalm 106			Gaelic Psalm 22		
Words	Verses	Z-A	Words	Verses	Z-A
3	7	2.64	4	11	11.44
4	30	31.83	5	43	42.61
5	70	67.81	6	36	36.51
6	45	48.41	7	14	13.57
7	23	18.90	8	4	3.62
8	4	5.64			

9	1	1.95			
a = 42.9799367 b = -13.1898584 c = 4.1792278E-014 R^2 = 0.9844			a = 58.4149 b = -17.4302 c = 2.49867889E-020 R^2 = 0.9992		

Gaelic Psalm 89			Gaelic Psalm 107		
Words	Verses	Z-A	Words	Verses	Z-A
3	3	2.63	3	4	1.28
4	31	31.54	4	16	13.62
5	70	69.13	5	44	46.71
6	50	51.53	6	49	45.27
7	23	21.09	7	17	21.15
8	6	6.50	8	9	6.80
9	1	2.20	9	1	2.25
a = 42.2759444 b = -12.911623 c = 6.43650511E-014 R^2 = 0.9978			a = 49.6690963 b = -14.6549996 c = 2.68463569E-017 R^2 = 0.9734		

Gaelic Psalm 37			Gaelic Psalm 104		
Words	Verses	Z-A	Words	Verses	Z-A
3	3	5.21	3	5	2.05
4	29	26.69	4	22	20.77
5	39	41.91	5	47	49.31
6	36	33.57	6	45	41.85
7	19	18.53	7	18	19.88
8	6	8.48	8	6	7.07
			9	1	2.56
a = 26.7255756 b = -8.22524685 c = 1.52898657E-008 R^2 = 0.9733			a = 40.3866493 b = -12.1446769 c = 1.3095497E-013 R^2 = 0.9848		

Gaelic Psalm 119			Gaelic Psalm 68		
Words	Verses	Z-A	Words	Verses	Z-A
3	18	18.03	3	9	4.67
4	124	121.96	4	27	28.15
5	185	189.15	5	48	48.04
6	140	135.11	6	39	39.47
7	63	62.66	7	26	21.57
8	18	22.87	8	3	9.53
9	2	7.62			
a = 30.3282497			a = 28.8103115		

b = -9.46292484 c = 5.26340048E-009 R^2 = 0.9967	b = -8.79483242 c = 2.68327415E-009 R^2 = 0.9438

Tables 3 and 4 provide a summary of the estimates for parameters *a* and *b*. (Parameter c is negligible in all cases and will be commented on in the next section.)

Table 3: Parameters a and b in the Welsh data

Psalm	a	b
18	23.84	-7.49
22	32.26	-10.08
37	29.04	-9.03
68	30.40	-9.74
78	29.97	-9.50
89	32.01	-9.91
104	25.50	-7.79
106	20.15	-6.40
107	22.15	-6.98
119	31.78	-9.88

Table 4: Parameters a and b in the Scottish Gaelic data

Psalm	a	b
18	52.93	-15.99
22	58.41	-17.43
37	26.73	-8.23
68	28.81	-8.79
78	39.33	-11.74
89	42.28	-12.91
104	40.39	-12.14
106	42.98	-13.19
107	49.67	-14.65
119	30.33	-9.46

4. Conclusion

In all but one case (namely the Welsh translation of psalm 106), an acceptable model fit was possible using the Zipf-Alekseev function. Even this one exception was only slightly below the R^2 cut-off of 0.8, with a value of 0.78. Most probably this is due to a boundary condition - e.g., corrections by editors. Overall, this experiment suggests that the Zipf-Alekseev function may well prove to be a generalizable model for verse lengths in both Welsh and Scottish Gaelic. It also provides further support for it as a cross-linguistic model for verse length.

Examining Tables 3 and 4, it is clear that parameter *a* is always a positive number here, in the approximate range [20, 60]. The values of parameter *a* tend, on the whole, to be larger for the Scottish Gaelic texts (minimum = 26.73, maximum = 58.41) than for the Welsh texts (minimum = 20.15, maximum = 32.26); seven out of the ten Gaelic texts have parameter *a* larger than the maximum for the Welsh texts. Parameter *b* is always a negative number in the approximate range [-18, -6]. The absolute values of parameter *b* tend again to be larger in the Gaelic texts than in the Welsh texts: the precise ranges, as negative numbers, are [-17.43, -8.23] for Gaelic and [-10.08, -6.4] for Welsh, with seven out of the ten Gaelic texts falling below the lower bound of the range for Welsh.

It is notable that parameter *c* is very small in all cases – indeed, one might say negligible, as it is far smaller than the two or three decimal places to which figures are typically rounded. If this continues to prove the case with other Welsh and Gaelic texts, it might then be more appropriate, in future, to attempt to model verse lengths in these languages using the simple power function: $y = ax^b$.

Modelling the relationship between parameters *a* and *b* for these texts shows that they are also linked by a power function $y = cx^d$ in both Welsh and Gaelic. The Welsh data give estimates of $c = -0.327$ and $d = 0.987$, with $R^2 = 0.992$. The Gaelic data give estimates of $c = -0.356$ and $d = 0.957$, with $R^2 = 0.997$.

Based on these few data alone, little more can be said about the estimated parameters from a functional perspective. Further work should attempt to apply these insights to other verse data from Welsh and Scottish Gaelic, in order to arrive at a full and explanatory synergetic model of verse lengths in these two languages. Patterns of parameter variation according to author, date, genre, etc., will need to be examined.

Acknowledgements and Data Access Statement

The Scottish Gaelic texts are from Text 188 - Sailm Dhaibhidh - courtesy of the Digital Archive of Scottish Gaelic project (www.dasg.ac.uk). The Welsh texts are from Morgan (2011).

References

Best, K.-H.. (2012a). How many words are in a verse? An exploration. In: Naumann, S., Grzybek, P., Vulanovic, R., Altmann, G., (eds.), *Synergetic linguistics. Text and language as dynamic systems:13-22*. Wien: Praesens.

Best, K.-H. (2012b). Zur Verslänge bei G. A. Buerger. *Glottometrics 23, 56-61*.

Best, K.-H. (2013). Zur Verslänge im Altisländischen. *Glottometrics 25, 22-29.*

Grotjahn, R. (1979). *Linguistische und statistische Methoden in Metrik und Textwissenschaft.* Bochum: Brockmeyer.

Koch, V. (2014). *Quantitative film studies: Regularities and interrelations exemplified by shot lengths in Soviet feature films*. 2 vols. Ph.D. Thesis, University of Graz.

Köhler, R. (1987). System theoretical linguistics. *Theoretical Linguistics 14, 241–257.*

MacTavish, D.C. (1934). *The Gaelic Psalms, 1694*. Being a reprint of the edition issued by the Synod of Argyll in that year. With an historical introduction by Duncan C. MacTavish. Lochgilphead: James M.S. Annan.

Mačutek, J. & Altmann, G. (2007). Discrete and continuous modelling in quantitative linguistics. *Journal of Quantitative Linguistics 14(1), 81-94.*

Morgan, A. (2011). *Astudiaeth o Salmau Cân (1621) Edmwnd Prys*. 2 vols. Ph.D. thesis, Aberystwyth University.

Muller, C. (1972). *Einführung in die Sprachstatistik*. München: Hueber.

Popescu, I.-I., Best, K.-H., & Altmann, G. (2014). *Unified modeling of length in language*. Lüdenscheid: RAM-Verlag.

Williams, G.A. (2004). The poetic debate of Edmwnd Prys and Wiliam Cynwal. *Renaissance Studies 18(1), 33-54.*