An analysis of leaf shapes for a selection of deciduous trees

David E. Rydeheard*

1 Introduction

In order to support a semantics of botanical descriptions, we have undertaken a sampling of field data to evaluate the semantic modelling. In this brief note, we report some of the results.

We have sampled 7 species of deciduous trees, each sample taken from an area around Manchester (UK), late in the year (October/November). Larger samples covering a greater geographic area could be assembled, but the purpose at present is limited to illustrating the kind of analysis that we can perform to assess the semantics.

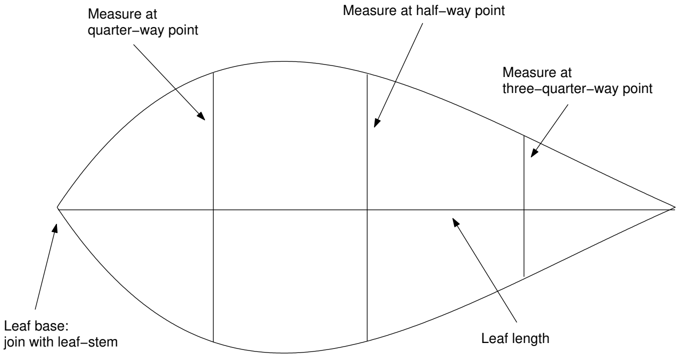

Figure 1: Diagram of leaf measurements.

To capture the range of shapes of the leaves, each leaf was measured from its base (at the join of the lamina with the petiole) to the end of the main lamina shape (this will not coincide with the leaf apex/tip, if this is elongated or truncated). The width of each leaf was measured at points along the length, one quarter way, half way, and three quarters way along from the base to the end of shape. See Figure 1 for a diagram depicting these measurements.

Let us look at the results for an example species: *Fagus sylvatica* (Beech) with leaves variously described as:

oval or obovate
oval to elliptical

*School of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, U.K. emails: david.rydeheard@manchester.ac.uk.

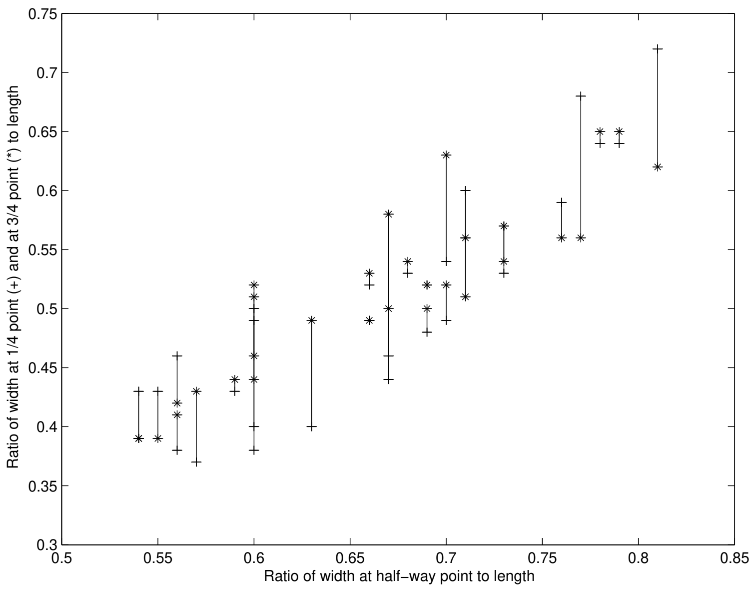

Figure 2: Ratios of leaf dimensions for a sample of *Fagus sylvatica* (Beech).

elliptical
oval
ovate to elliptic.

In Figure 2, for a sample of *Fagus sylvatica* leaves, we plot the ratio of quarter-way width to length (marked with $+$) and the ratio of the three-quarter-way width to the length (marked with $*$) (both along the y-axis) against the ratio of the half-way width to the length (along the x-axis).

What do we observe from this plot? We now describe features of real leaves and how they are reflected in leaf descriptions:

1. Firstly, notice that, in this sample, the ratio of leaf width at the half-way point to the leaf length (the figures on the x-axis) varies from 0.54 to 0.81 i.e. the leaves are 1.23 to 1.85 times as long as they are wide at the half-way point. Thus some leaves are almost 'circular' and others are longer and narrower, with a fairly even spread between these extremes.

2. Some leaves are almost symmetric about the half-way point (i.e. approach quadripartite symmetry) – these are the points where the quarter-way ratio (+) and three-quarter-way ratio (*) coincide or are close and such leaves are *elliptic*. Those with a larger quarter-way ratio are *elliptic-ovate*, whilst those with a larger three-quarter-way ratio are *elliptic-obovate*. We see that, in this sample, there is a fairly even spread between these three forms of shape, with some extreme cases too. Notice also that no leaves in this sample are *ovate* (quarter-way ratio is larger than half-way ratio) or are *obovate*.

3. We expect that as the half-way ratio increases, so do the other two measures (i.e. the leaf becomes 'fatter'). What does this dependency look like? Both sets of

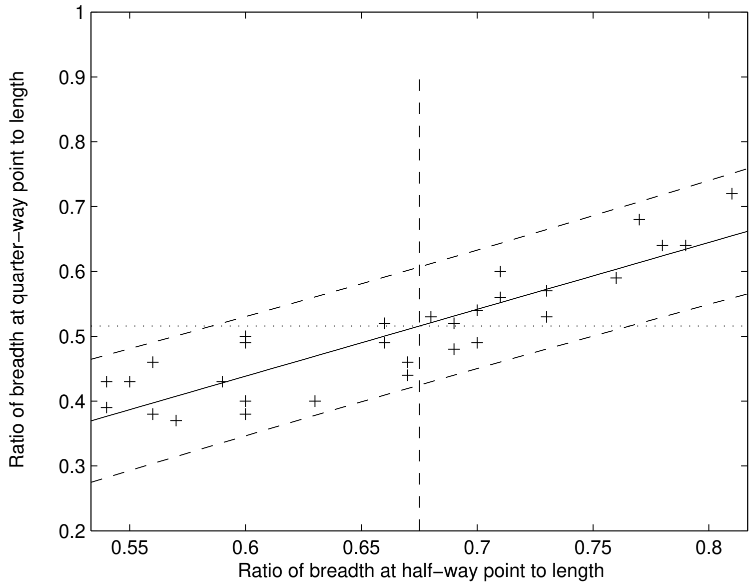

Figure 3: A straight line fit for leaves of *Fagus sylvatica* (Beech).

data, the quarter-way ratios and the three-quarter-way ratios, approximate straight (but different) lines. The approximation to straight lines is the geodesic for the Euclidean metric in the first of the models above, saying that the proportionate increase in each of the ratios is constant when moving between shapes. How good is this approximation for this sample? In Figure 3, we plot the straight line that is the best least-squares fit using a linear regression model for the quarter-way ratio. The line has gradient 0.95, that is an increment in the half-way width is approximately matched by the same increment in the quarter-way width.

The 95% confidence limits are indicated on Figure 3 as the upper and lower curves. These are almost constant at approx. 0.09 above and below the fitted line and there is a fairly uniform distribution of data points around the line.

4. Variation with size: It may have occurred to the reader that leaf shape may vary with size, even for mature leaves. An analysis of leaf-ratios to actual leaf length for this species reveals no apparent trend in the shape as the leaf gets larger. However, in some species there does appear to be a variation of shape with size. This is a form of variation which is rarely mentioned in floras, except where there is a gross difference between the shapes of smaller and larger leaves (or other causes of variation of leaf shape, e.g. species with juvenile foliage). Thus with this form of variation, descriptions might be of the form *small A to large B* where A and B are leaf shape expressions. This may be modelled using geodesics in product spaces. The two spaces in this example are the linear measurement of 'size' and the space of leaf shapes.

We continue the analysis of leaf shapes in the text by looking briefly at the data gathered from samples of 6 other species collected in the same area at the same time. In this, we will see different sorts of shapes and of variation between shapes so as to give us a fuller picture of how leaf shape descriptions match actual leaves.

In Figure 4, we display a sample of leaves of *Prunus avium* (Wild cherry). This is variously described as:

obovate to elliptic
oval
oblong-ovate or ovate
oval to oblong
oblong.

We plot the same dimensions as for Beech leaves in the text, namely the ratios of the width at the quarter-way point (+) and the three-quarter-way point (*) against the ratio of the width at the half-way point to the length.

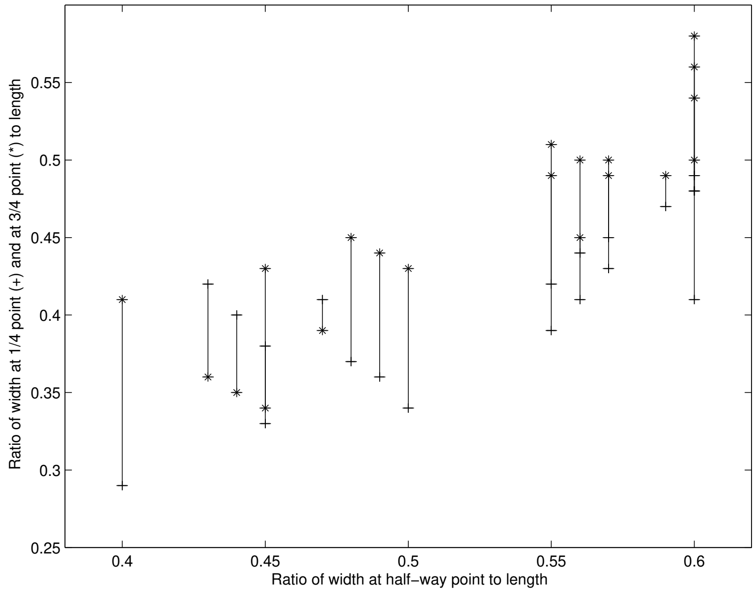

Figure 4: Sample ratios of leaf dimensions for *Prunus avium* (Wild cherry).

We see that the leaves in this sample vary between 1.7 to 2.5 as long as wide at the half-way point. There is a great preponderance of leaves wider at the three-quarter-way point that the quarter-way point indicating nearer an obovate than ovate shape. Both sets of points lie roughly diagonally. Straight line plots (using least-squares fit with linear regression) show two lines of different gradient: for quarter-way points, the gradient is 0.65 (and 95% confidence bounds are at 0.075), and for three-quarter-way points, the gradient is 0.9 (with 95% confidence bounds at 0.07).

In Figures 5, 6, 7, 8 and 9, we display the same data for a range of other deciduous trees: *Prunus spinosa* (Blackthorn), *Euonymus europaeus* (Spindle), *Nothofagus procera* (Rauli), *Maclura pomifera* (Osage orange), and *Ulmus glabra* (Wych elm). Looking at these plots, it is clear that the form of leaf variation itself varies from species to species, and the extent to which the textual descriptions (included in the captions to the Figures) captures the variation itself is variable.

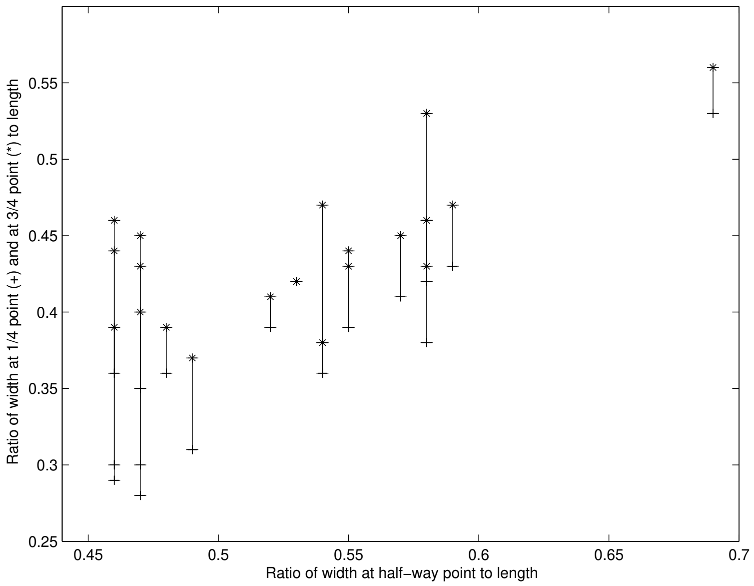

Figure 5: Sample ratios of leaf dimensions for *Prunus spinosa* (Blackthorn). This is variously described as: obovate to oblanceolate; oval, broadest above middle; and, oval.

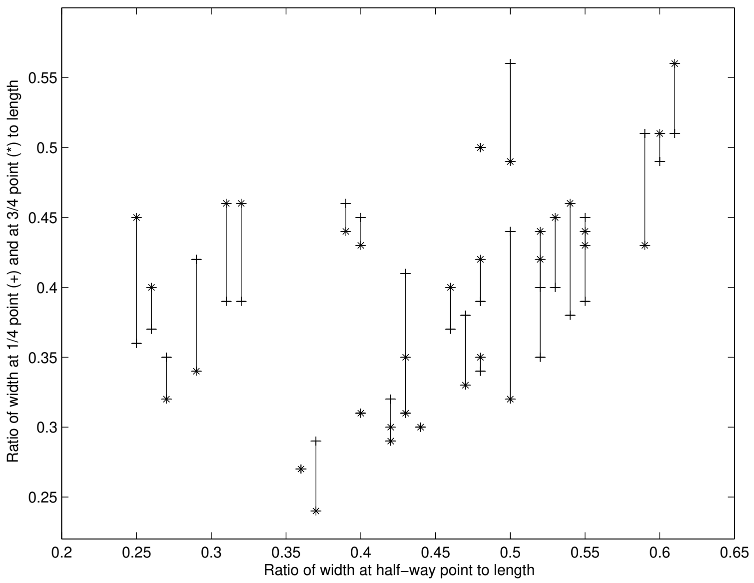

Figure 6: Sample ratios of leaf dimensions for *Euonymus europaeus* (Spindle). This is variously described as: elliptical to oval; lanceolate to elliptical; lanceolate; and, elliptic to narrowly so or narrowly obovate.

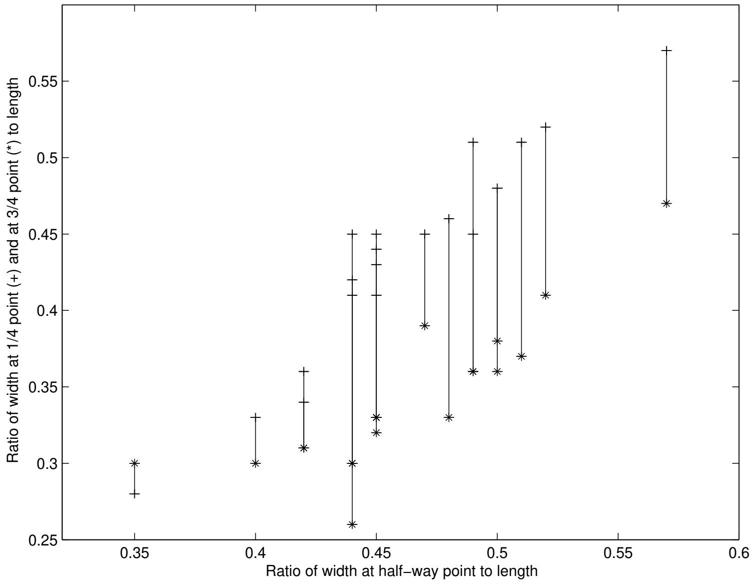

Figure 7: Sample ratios of leaf dimensions for *Nothofagus procera* (Rauli: Argentina, Chile). This is variously described as: ovate-lanceolate or oblong-ovate; oblong to ovate-lanceolate; ovate-lanceolate to trullate.

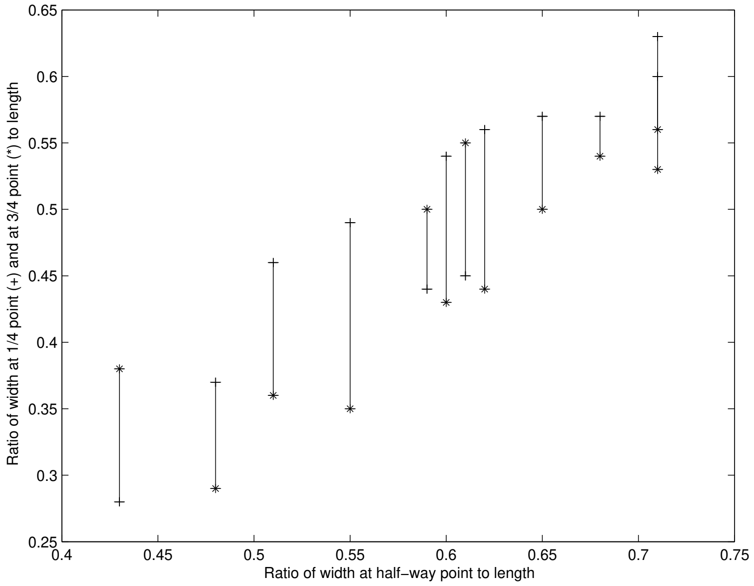

Figure 8: Sample ratios of leaf dimensions for *Maclura pomifera* (Osage orange: Central USA). This is variously described as: ovate, short acuminate, or lanceolate, long-acuminate, mixed sizes in whorls; lanceolate, oblong, ovate; ovate; ovate or ovate-lanceolate.

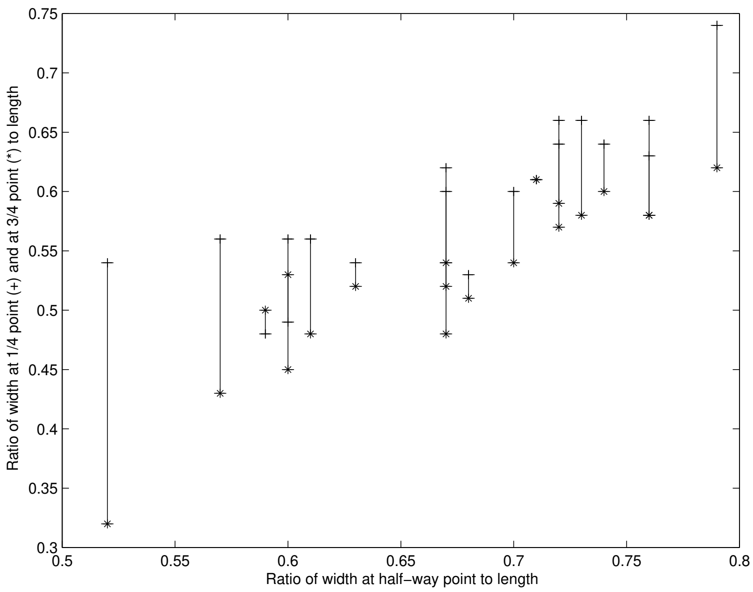

Figure 9: Sample ratios of leaf dimensions for *Ulmus glabra* (Wych elm). This is variously described as: obovate; broad lanceolate; rounded to elliptical; and, (two variants) broadly obovate/narrowly obovate.

What is also clear from this preliminary analysis is the lack of unanimity in the textual descriptions, even for this simple leaf shape, and the corresponding variation in nature. Notice that though the samples chosen are small, increasing their size can only increase variation present (though it may lead to more clustering around points or paths). For the multiple parallel descriptions, a key idea is that of 'integration' – combining the descriptions into a single expression which suitably captures the breadth of meaning without redundancy and detecting any possible inconsistency. Integration requires semantic information, and we have already begun to build automatic tools based on the semantics. Of course, an integrated description can also be tested against nature to assess both the modelling and the integration techniques. As for the range of variation present even in this simple leaf shape, it is clear that our language for describing shape struggles not only in describing complex shapes but also in describing the complex variation present in nature.