

# Units and Constituency in Prosodic Analysis: A Quantitative Assessment

Andrew Wilson

## Abstract

Drawing on methods from quantitative linguistics, this paper tests the hypothesis that the intonation unit is a valid language construct whose immediate constituent is the foot (and whose own immediate constituent is the syllable). If the hypothesis is true, then the lengths of intonation units, measured in feet, should abide by a regular and parsimonious discrete probability distribution, and the immediate constituency relationship between feet and intonation units should be further demonstrable by successfully fitting the Menzerath-Altmann equation with a negative exponent. However, out of sixteen texts from the Aix-MARSEC database, only six share a common probability distribution and only eight exhibit a tolerable fit of the Menzerath-Altmann equation. A failure rate of  $\geq 50\%$  in both cases casts doubt on the validity of the hypothesis.

**Key words:** prosody; intonation; constituency; synergetic linguistics; quantitative linguistics

## 1. Introduction

### 1.1 Aims

The approach to prosodic analysis that is commonly called the 'British' tradition recognizes four levels of units: phonemes, syllables, feet, and intonation units (the latter sometimes called 'tone groups' instead). In some versions of this approach, notably in Halliday's (1967) systemic functional theory, these levels are also deemed more specifically to stand in an immediate constituency hierarchy to one another, such that syllables are made up of phonemes, feet are made up of syllables, and intonation units are made up of feet. However, a number of these foundational claims are controversial, and have been for some time. For instance, Crystal (1969, p. 362), in a detailed critical review of Halliday (1967), raised at least the possibility that further intermediate units might exist - e.g., in between the intonation unit and the foot - and also questioned the operational definition of the foot itself. Although Crystal was writing more than forty years ago, the issues have not gone away: on the one hand, the model continues to be used in practice (e.g. Halliday & Greaves, 2008), whilst, on the other hand, debates about its adequacy continue. For instance, Bouzon and Hirst (2004) have compared the Hallidayan foot-based model unfavourably with an alternative model of English speech rhythm proposed by Jassem (1952). This paper aims to take the debate a step further by providing a quantitative test of the hypothesis that the intonation unit is a valid language construct whose immediate constituent is the foot (and whose own immediate constituent is the syllable).

### 1.2 Units of prosodic analysis

Halliday's (1967) definition of the foot is taken over from Abercrombie (1964). He defines the foot as extending from a stressed syllable or intonation group boundary to the following stressed syllable, without, however, incorporating the second stressed syllable. In so far as the foot, under this definition, can extend merely from one stressed syllable to the next, the definition is identical to the long-standing

notion of a 'rhythmical unit', laid out at the very beginning of the 20th century by the German psychologist Karl Marbe (1904; see also Best, 2002; 2005). What is controversial, however, in Abercrombie's definition, is the addition of the intonation unit boundary as a foot delimiter. This introduces the possibility of encountering a foot that contains no stressed syllable, whereas most traditional definitions of feet (e.g. in poetics) require them to include at least one stressed syllable. In the Abercrombie-based framework, those feet which lack a stressed syllable are sometimes said to have a 'silent ictus' in place of a realized stressed syllable. Crystal (1969, p. 382) refers to this notion of the silent ictus as being "of very dubious status".

The problem of units within the 'British' scheme of prosodic analysis does not end with the definition of feet. Perhaps even more controversial is the notion of the intonation unit. Whilst almost everyone recognizes that higher units of phrasing exist within speech, the intonation unit remains remarkably poorly defined in operational terms. Technically, in the 'British' tradition, the delimitation of an intonation unit hinges on the recognition of the nuclear syllable (also known as the nucleus or the tonic). This is the syllable on which the perceptually most salient pitch movement occurs, and it is the only obligatory element in an intonation unit. (Other optional syllables in the intonation unit make up the pre-tonic and the post-tonic or tail.) What is problematic about this definition of the intonation unit is the set of criteria - or, rather, a lack of one - for identifying the nuclear tone, upon which everything else depends. As has just been said, intonation units are defined formally by the presence of this nuclear tone, yet Cruttenden (1969, p. 310) points out that, in Halliday (1967), "there is some implication that we should work in the opposite direction, that is firstly identify tone groups, then the tonic, which is 'obligatory in the tone group'". In other words, Halliday appears to be suggesting that acoustic or linguistic criteria other than the nuclear tone are relevant - and perhaps even more important - for defining the intonation unit. This goes against the theoretical definition but is certainly a widespread practice in prosodic annotation. Knowles (1991), for instance, places division into intonation units before the identification of accent and stress in his 'recipe' for prosodic annotation, and he furthermore lists the auditory/acoustic discontinuities that may be suggestive of intonation unit boundaries: temporal

discontinuities (i.e. pauses), pitch discontinuities, and segmental discontinuities (e.g. the absence of assimilation, elision, etc.). Especially in the context of speech synthesis research, many attempts to predict intonation group boundaries have also drawn on grammatical information and punctuation. More recent work has additionally proposed incorporating other non-traditional cues - for instance, recognizing a link between prosodic boundaries and the presence of words that include diphthongs or triphthongs (Brierley & Atwell, 2011).

In sum, however, in whatever way we approach it, the intonation unit remains a rather nebulous construct. We have an intuition that something like it exists but we cannot agree exactly on how to define and recognize it operationally. For instance, nuclear tones are perceptual: whilst we can locate pitch movements objectively by looking at fundamental frequency traces, deciding which of these many movements count as 'nuclear' is ultimately a subjective decision. The trouble is that people's subjective judgements can differ. Ideally, we need some kind of formal procedure which can assist us in deciding whether intonation units, when identified operationally under one specification or another, are actually 'legitimate' linguistic constructs, with Abercrombie's (1964) feet as their immediate constituents. I would argue that (synergetic) quantitative linguistics can provide the necessary tests to address this problem.

### 1.3 The contribution of quantitative linguistics

Taking its inspiration from the quantitative physical sciences (Altmann & Meyer, 2005; Altmann, 2008), quantitative linguistics starts out from the position that "everything abides by laws" (Bunge, 1977, p. 16) and attempts to integrate these laws into a view of language as a dynamic self-organizing system in which form and function are interdependent. The currently dominant paradigm is that of Köhler's systems-theoretic linguistics (Köhler, 1986; Köhler, 1987), although other viewpoints are also possible. Köhler (1987, p. 243) stresses that laws are the only basis for a linguistic theory that can explain why language, as a human phenomenon, behaves as it does:

"Explanation in the scientific sense of the word is only possible by means of laws. Grammars and rules are not invariant: they are neither universally valid for all languages, nor are the structures which they describe constant over time (languages change). Therefore, rules cannot explain anything - on the contrary: they need *to be explained*."

In other words, linguistic laws operate on a different level to the statements of rules, grammars, etc., of other linguistic theories: they are fundamental regularities that hold true for all languages at all times and which explain how language, in general, operates. If a provisional law fails on this latter criterion, then it is no longer a valid language law in the strict sense of the term.

The laws recognized by quantitative linguistics are of three kinds: functional, distributional, and developmental (Köhler & Altmann, 2005, p. 34). In this paper, we shall be concerned with laws of the first two types.

A distributional law, on the one hand, is a (most often univariate) stochastic law and establishes a theoretical probability distribution by which the frequencies of particular linguistic units, or their properties, abide. For instance, it is very well established that word frequencies, when sorted into rank order, abide by the 1-displaced Zeta distribution (more colloquially known as "Zipf's Law") (Popescu et al., 2009). Many other examples of such units and their distributions are attested in the literature (see, e.g., Best, 2001, 2011; Köhler, Altmann & Piotrowski, 2005); of particular relevance in the present context is Best's (2002; 2005) finding that the lengths of the rhythmical units defined by Karl Marbe (see above) abide, in German prose, by the 1-displaced hyper-Poisson distribution.<sup>1</sup> For the present purpose, however, what is important is not, in the first place, the particular distribution which is established for a unit or property but rather the fact that the unit or property abides by a distribution at all. If a linguistic unit is 'legitimate', it will exhibit a regular probability distribution - i.e., in repeated experiments, the same

---

<sup>1</sup> Nearly all linguistic distributional laws entail 1-displaced distributions, because one cannot normally have words with zero frequency or units with zero length.

distribution will hold true, with only a negligible proportion of 'failures'; conversely, 'non-legitimate' units will behave randomly: they will either fail to exhibit a theoretical distribution at all, on multiple occasions, or they will oscillate between different probability distributions without any clear boundary conditions to account for the variation.

There is a further corollary to this observation. When measuring the lengths of linguistic units in such experiments, the measurement must normally be performed in terms of the immediate constituent.

Altmann and Meyer (2005, pp. 43-44) illustrate this in relation to sentence length. They note that "if sentence length is measured in clauses, its distribution abides by the negative binomial" (which has two parameters) but "if measured in words, a supplementary parameter must be introduced", leading to the hyper-Pascal distribution, which has three parameters. Putting it bluntly, they comment that "the more we descend on the ladder of immediate units the more chaotic is the result" - for example, if sentence lengths are measured in phonemes (e.g. Clayman, 1981) rather than in clauses. In other words, the regular occurrence of a probability distribution with a minimal number of parameters (usually just one or two) is only to be expected if the measurements are undertaken using the immediate constituent as the unit of measurement. A distribution with a minimal number of parameters is the desideratum in quantitative linguistics because the parameters should, at least in the long run, be functionally interpretable and not merely mathematical artifacts.<sup>2</sup>

In contrast to a distributional law, a functional law is typically bivariate and describes the relationship between one linguistic unit or property and another. In quantitative linguistics, there is a clearly defined functional law governing constituency relations, regardless of whether they are phonological, morphological, syntactic, or textual in nature. This is known as the Menzerath-Altmann law, and it states (when expressed in qualitative terms) that the length of a linguistic unit is a function of the number of its constituents. The law derives from the observations of Paul Menzerath (1954) on German morphology,

---

<sup>2</sup> For example, in some studies of word-length frequencies, systematic variation in the parameters of the same probability distribution can be linked to text type (Antić et al., 2006).

which were later formalized mathematically in a series of publications by one of the founding fathers of contemporary quantitative linguistics, Gabriel Altmann (e.g. Altmann, 1980; Altmann & Schwibbe, 1989).

The simple qualitative statement of the Menzerath-Altmann law belies the mathematical detail which enables the law to be used as a test for immediate constituency. Mathematically, the Menzerath-Altmann law is given by:

$$y = C * x^a * \exp(bx)$$

where  $y$  is the mean length of the constituent unit and  $x$  is the length of the higher-level construct to which it belongs. (For instance,  $y$  might be the mean length of feet within intonation units, measured in syllables, and  $x$  might be the length of intonation units, measured in feet.)  $C$  is a constant, and  $a$  and  $b$  are parameters, all of which can be estimated using customizable curve-fitting software. When  $b = 0$ , the equation simplifies to the commonly encountered  $y = C * x^a$ ; however, the full equation is normally required when the equation involves units smaller than the clause (Andres, 2010).

According to Altmann and Meyer (2005, p. 43), who used  $y = C * x^a$ , "a unit is an immediate component of a higher unit only if they are bonded by Menzerath's law with a *negative exponent*" - i.e., with a negative parameter  $a$  in the equation  $y = C * x^a$ . In other words, if the equation cannot be fitted adequately, or a positive value of  $a$  is obtained, then we are justified in saying that there is no immediate constituency relationship between the unit and its hypothesized constituent unit. In a case where there is a good quality of fit, but a positive  $a$  is obtained, Altmann and Meyer (2005) show that there must be another level of constituency intervening between the two units entered into the equation. When the full equation is used, as here, the corresponding requirement becomes simply that at least one of  $a$  and  $b$  is negative (Altmann, personal communication). All of these observations have been tested on multiple occasions with a wide variety of linguistic units and typologically differing languages (e.g. Cramer, 2005).

Putting everything together, we should expect that, if intonation units, as operationalized in an annotated speech database, are 'legitimate' language constructs, and feet are their proper immediate constituents, then (1) the lengths of the intonation units, measured in feet, will abide by a regular discrete probability distribution (ideally with no more than two parameters), and (2) the immediate constituency relationship between feet and intonation units will be further demonstrable by successfully fitting the Menzerath-Altmann equation with a negative exponent  $a$  and/or  $b$ .

## 2. Data and Method

The data for this experiment were selected from version 2 of the Aix-MARSEC corpus (Roach, Knowles, Varadi & Arnfield, 1993; Auran, Bouzon & Hirst, 2004). This is a corpus of naturally occurring formal spoken British English, dating from the 1980s. It consists largely of written-to-be-spoken material, much of it from radio broadcasts, and covers a range of genres: news broadcasts, academic lectures, charity appeals, story and poetry readings, etc. Speakers were selected for inclusion in the corpus on the basis that their speech approximated to British Received Pronunciation (RP).

The corpus has been richly annotated with phonetic and phonological information (Knowles, 1994). One of the most important annotation layers is that of intonation. These annotations were undertaken on an auditory-impressionistic basis by two trained phoneticians with roots in the 'British' tradition of prosodic analysis: Gerry Knowles (GOK) and Briony Williams (BJW). Each phonetician annotated roughly half of the corpus, inserting tonetic stress markers and intonation unit boundaries according to a version of the 'British' system of prosodic analysis (Knowles, Wichmann & Alderson, 1996). The annotation was made using unpunctuated orthographic transcripts to ensure that the phoneticians' auditory judgements of phrasing were not biased by the presence of punctuation marks. Roughly 10% of the corpus was transcribed independently by *both* phoneticians, resulting in the so-called "overlap passages". Previous



analyses of these passages have demonstrated notable differences between the judgements of the two phoneticians (Knowles, 1991; Knowles, 1992).

For this experiment, sixteen texts were selected from Aix-MARSEC. This was the maximum possible number of texts available once the following objective criteria had been applied:

1. Texts involving multiple speakers were excluded. Each person has their own speaking style, and the mixing of persons disturbs the homogeneity of the spoken text. Heterogeneous texts always entail at least some degree of stratification according to speaker (Altmann, 1992; Ziegler & Altmann, 2003; Knight, 2013) and have consequently proven to be problematic for quantitative modelling.
2. For similar reasons, texts whose annotation was split between both phoneticians were excluded. These were mostly longer texts in the genre of the academic lecture. This is a separate issue from the overlap passages referred to above, as only a single phonetician worked on each of the text parts in this case. Since we know already (from Knowles's 1991 study) that the annotators' practices showed some differences, a mixing of annotators would again disturb the homogeneity of any patterning within the text.
3. Incomplete texts were also excluded. On the whole, it has become the standard practice in quantitative linguistics to use only complete texts (Altmann, 1992). Whilst patterns at some linguistic levels (like grapheme and phoneme counts) appear to be unaffected by sampling from longer texts, and even by mixing of texts (e.g. Kelih, 2006), others only exhibit stable patterns when complete homogeneous texts are analysed. The experiments by Best (2002) on Marbe's rhythmical units, which have some relationship to the units studied in this paper, showed that regular probability distributions of rhythmical unit lengths tended only to hold for complete texts and not for samples from texts.
4. Finally, poetry recitals were excluded on the grounds of both length and their inherent metrical patterns, which are very different from those of ordinary prosaic speech (Barney, 1999).

The resulting sample of texts thus consisted of eight texts annotated by GOK and eight annotated by BJW. These texts belonged to three different genre categories, in the same proportions for both annotators: six current affairs stories from radio broadcasts (texts A1 to A12); one radio charity appeal (texts K1 and K2); and one presentation speech for an honorary degree (texts M5 and M6). GOK annotated texts A2, A4, A5, A7, A9, A11, K1, and M6, whilst BJW annotated texts A1, A3, A6, A8, A10, A12, K2, and M5. Further details of the texts can be found in Taylor and Knowles (1988).

From each text, individually, the following information was extracted: (1) the length of each intonation unit in terms of feet and (2) the length of each intonation unit in terms of syllables. From these, it was also possible to calculate the mean foot length (in terms of syllables) within an intonation unit. Two different statistical analyses were then undertaken.

First, the frequencies of each length class of intonation units (i.e. the number of one-foot intonation units, the number of two-foot intonation units, etc.) were counted for each text. These frequency tabulations were then analysed using the Altmann-Fitter software (Köhler & Altmann, 1997). This program iteratively fits 205 discrete univariate probability distributions (Wimmer & Altmann, 1999) in order to find the best-fit distribution for the data. Two measures of goodness of fit are provided for each distribution tested: (1) the conventional Pearson chi-squared test and (2) the discrepancy coefficient  $C$ . The latter is given by  $\chi^2/N$ , where  $N$  (in this case) is the number of intonation units. The chi-squared test *per se* has been known for a long time to be problematic: in particular, the value of chi-square increases linearly with sample size, so that larger samples readily show divergence from the theoretical distribution if a conventional criterion of, e.g.,  $p > 0.05$  is used (Popescu, Mačutek & Altmann, 2009, pp. 15-16; Rietveld, van Hout & Ernestus, 2004). In quantitative linguistics, it has therefore become increasingly accepted to use  $C$  as the main measure of goodness of fit (Grzybek, 2006, p. 23). A value of  $C$  less than 0.02 is normally taken to indicate an acceptable fit, with values less than 0.01 indicating a very good fit.

Second, the full version of the Menzerath-Altmann equation was fitted to the data from each text. This was undertaken using the `nls` function within the R environment for statistical computing (Ihaka & Gentleman, 1996) and the R code provided by Andres et al. (2012). Goodness of fit was established using the determination coefficient  $R^2$ , which is given by:

$$R^2 = 1 - ( \text{sum}((\text{obs}-\text{pred})^2) / \text{sum}((\text{obs}-\text{mean}(\text{obs}))^2) )$$

where `obs` = the observed value and `pred` = the predicted value.

In quantitative linguistics, a fit is generally considered good if  $R^2$  is equal to, or greater than, 0.9 (Popescu, Mačutek & Altmann, 2009, p. 16). A fit with  $0.9 > R^2 > 0.7$  is tolerable, but not good.

### 3. Results

Tables 1 and 2 show the results of fitting two discrete probability distributions to the frequencies of intonation unit length classes measured in feet. Using the explorative Altmann-Fitter software, four distributions initially suggested themselves as possible candidate models across the sixteen texts, in so far as they each obtained a good quality of fit on more than just a couple of texts: the hyper-Poisson distribution, the Dacey-Poisson distribution, the Generalized Dacey-Poisson distribution, and the Dacey-Negative-Binomial distribution. Of these, the first pair have two parameters each and the second pair three parameters. The generalized Dacey-Poisson and the Dacey-Negative-Binomial distributions are therefore disregarded here on the grounds of parsimony, as outlined earlier. Table 1 shows the results of fitting the hyper-Poisson distribution and Table 2 shows the results of fitting the Dacey-Poisson distribution. Since we exclude the possibility of intonation units with a length of less than one foot, all distributions are considered in their 1-displaced form.

The hyper-Poisson distribution succeeds for six out of the sixteen texts: four out of eight for GOK's annotations and two out of eight for BJW's annotations. The Dacey-Poisson distribution also succeeds for six texts: five out of eight for GOK's annotations and one out of eight for BJW's annotations.<sup>3</sup>

Table 3 shows the results of fitting the Menzerath-Altmann equation. Although only one text failed to show a negative parameter  $a$  and/or  $b$ , it will be seen that a good fit of the equation ( $R^2 > 0.9$ ) could only be achieved for four of the sixteen texts and a tolerably acceptable fit ( $R^2 > 0.7$ ) for just another four. That means that the equation succeeded for only half of the total sample. The good and tolerable fits are split equally between the two annotators, each having two good and two tolerable fits.

Table 1. Fitting of the 1-displaced hyper-Poisson distribution.

| Text | $\chi^2$ | $P(\chi^2)$ | C      | DF |
|------|----------|-------------|--------|----|
| A2   | 1.44     | 0.8366      | 0.0085 | 4  |
| A4   | 1.34     | 0.7200      | 0.0045 | 3  |
| A5   | 5.44     | 0.2453      | 0.0248 | 4  |
| A7   | 1.98     | 0.7401      | 0.0116 | 4  |
| A9   | 5.30     | 0.2575      | 0.0248 | 4  |
| A11  | 2.46     | 0.6523      | 0.0127 | 4  |
| K1   | 6.04     | 0.1963      | 0.0271 | 4  |
| M6   | 6.19     | 0.1857      | 0.0236 | 4  |
| A1   | 4.17     | 0.5255      | 0.0254 | 5  |
| A3   | 3.54     | 0.6175      | 0.0244 | 5  |
| A6   | 1.27     | 0.9385      | 0.0080 | 5  |
| A8   | 3.48     | 0.7465      | 0.0281 | 6  |

<sup>3</sup> Although the detailed results are not presented here on the grounds of parsimony (parameter numbers), it is worth noting that the Generalized Dacey-Poisson and Dacey-Negative-Binomial distributions performed little better: in both cases, a satisfactory fit could not be achieved for nine out of the sixteen texts.

|     |       |        |        |   |
|-----|-------|--------|--------|---|
| A10 | 31.44 | 0.0000 | 0.2082 | 6 |
| A12 | 5.34  | 0.2543 | 0.0456 | 4 |
| K2  | 7.25  | 0.2026 | 0.0522 | 5 |
| M5  | 2.11  | 0.9091 | 0.0172 | 6 |

Table 2. Fitting of the 1-displaced Dacey-Poisson distribution.

| Text | $\chi^2$ | $P(\chi^2)$ | C      | DF |
|------|----------|-------------|--------|----|
| A2   | 1.45     | 0.8351      | 0.0086 | 4  |
| A4   | 1.29     | 0.7309      | 0.0044 | 3  |
| A5   | 5.51     | 0.2390      | 0.0252 | 4  |
| A7   | 2.38     | 0.6653      | 0.0139 | 4  |
| A9   | 4.10     | 0.3923      | 0.0192 | 4  |
| A11  | 2.77     | 0.5964      | 0.0143 | 4  |
| K1   | 7.19     | 0.0660      | 0.0322 | 3  |
| M6   | 5.96     | 0.2020      | 0.0228 | 4  |
| A1   | 4.63     | 0.4632      | 0.0282 | 5  |
| A3   | 3.89     | 0.5660      | 0.0268 | 5  |
| A6   | 1.43     | 0.9205      | 0.0091 | 5  |
| A8   | 3.73     | 0.7137      | 0.0300 | 6  |
| A10  | 35.92    | 0.0000      | 0.2379 | 6  |
| A12  | 5.25     | 0.2625      | 0.0449 | 4  |
| K2   | 11.91    | 0.0361      | 0.0856 | 5  |
| M5   | 3.45     | 0.7508      | 0.0280 | 6  |

Table 3. Fitting of the Menzerath-Altmann equation.

| Text | C      | a       | b       | R <sup>2</sup> |
|------|--------|---------|---------|----------------|
| A2   | 2.2396 | -0.3935 | 0.1055  | 0.7075         |
| A4   | 2.3046 | -0.1194 | -0.0487 | 0.8461         |
| A5   | 1.8318 | -0.2196 | 0.0787  | 0.4128         |
| A7   | 2.0606 | -0.2947 | 0.0824  | 0.5361         |
| A9   | 2.0115 | 0.0005  | -0.0015 | 0.0035         |
| A11  | 2.4873 | -0.5759 | 0.1329  | 0.9218         |
| K1   | 2.0038 | 0.1554  | -0.0709 | 0.6344         |
| M6   | 2.2758 | -0.5750 | 0.1545  | 0.9070         |
| A1   | 1.7480 | 0.5600  | -0.1840 | 0.9393         |
| A3   | 2.5749 | -0.5183 | 0.0942  | 0.9238         |
| A6   | 2.4629 | -0.3346 | 0.0647  | 0.7227         |
| A8   | 1.6617 | 0.5515  | -0.1292 | 0.5168         |
| A10  | 2.0485 | -0.0782 | 0.0258  | 0.1061         |

|     |        |        |         |        |
|-----|--------|--------|---------|--------|
| A12 | 1.8749 | 0.3360 | -0.1203 | 0.7958 |
| K2  | 1.9393 | 0.1769 | -0.0464 | 0.5094 |
| M5  | 1.9768 | 0.0885 | 0.0079  | 0.5521 |

#### 4. Discussion

The fitting of discrete probability distributions to the frequencies of intonation unit length classes, measured in feet, showed mixed fortunes. Depending on which of the two final candidate distributions is chosen (and there is not a huge amount to choose between them in terms of how many texts show a good fit), four or five out of GOK's eight texts (i.e. roughly 50%) succeeded in showing lawful behaviour in relation to this property. Nevertheless, this still means that a shared distribution could not be fitted, at the accepted levels of  $C$ , to the other 50% or so of the texts. Whilst it is not uncommon to encounter a few texts to which it proves impossible to fit a distribution that has otherwise succeeded in an experiment or series of experiments, a failure rate of 50% seems unacceptable. Hence, the search for an appropriate regular probability distribution in relation to GOK's texts has to be considered unsuccessful overall. With BJW's texts, the fitting of the probability distributions was even less successful: the hyper-Poisson distribution could be fitted to only two texts, and the Dacey-Poisson distribution to only one. In total, taking both annotators together, the success rate is a maximum of just 6 out of 16 texts, or 37.5%.

Any limited optimism encouraged by these results evaporates when the Menzerath-Altmann equation is examined. This could be fitted tolerably in only 50% of the cases, with only four texts actually showing a quality of fit rated as "good".

Taken together, the results of these two analyses therefore strongly suggest one of two things: either (1) the foot - at least in Abercrombie's (1964) definition - is, in most cases, not the correct immediate constituent of intonation units or (2) *both* annotators have failed to delimit the intonation units correctly in many of their texts. A third possibility, loosely connected to the latter, is that the intonation unit, as defined in the 'British' prosodic tradition, is actually not a legitimate language construct at all, but this cannot be inferred directly from the results of the present experiment, as all the calculations are predicated upon a hypothesized constituency relationship between feet and intonation units, as laid down in, e.g., Halliday's (1967) theory.

In relation to the first possibility - that Abercrombie's (1964) feet are not the correct immediate constituents of intonation units - it might alternatively, at first glance, seem to be of value to try applying Marbe's (1904) rhythmical units to English texts, since these have so far shown lawful behaviour in analyses of German texts (e.g. Best, 2002, 2005) and, as stated earlier, quantitative linguistics seeks to identify regularities that are valid across all languages. However, Marbe's units are not entirely unproblematic, especially when one tries to pair them with higher-level prosodic units. Operationally, such a link is problematic because we could not possibly align the boundaries of Marbe's rhythmical units with the boundaries of intonation units in a relationship of constituency. Because they exclude the intonation unit boundary as a delimiter, Marbe's units must sometimes cross over two intonation units when spanning the gap from one stressed syllable to the next. This being the case, they cannot form constituents of those intonation units. The only way of allowing for constituency would be to modify the definition of the rhythmical unit so that it cannot cross intonation unit boundaries, which would leave us again exactly with Abercrombie's (1964) definition of a foot. Thus, whilst Marbe's rhythmical units might be worthy of investigation in themselves - i.e., as stand-alone units - they cannot be incorporated hierarchically into a theory that also includes intonation units.

Other possible relationships between prosodic units have been proposed in the literature. For instance,



using the same database as the present study, Bouzon and Hirst (2004) investigated Jassem's (1952) approach to English speech rhythm as an alternative to the Hallidayan theory and considered it superior. Perhaps the most distinctive aspect of Jassem's theory is that he de-couples rhythm from intonation, whereas the traditional 'British' approach to prosody (including Halliday) conflates the two in a single hierarchical framework. However, Bouzon and Hirst did not apply the theoretical approach or text-filtering criteria from synergetic quantitative linguistics employed in the present work. It should, furthermore, be noted that the units which Jassem (1952) proposed (called the anacrusis and the narrow rhythm unit) do not stand in an immediate constituency relationship to intonation units, so - just as with Marbe's (1904) rhythmical units - it is not possible to examine Jassem's model at this level of phrasing within the framework of the Menzerath-Altmann law. Nevertheless, it would be possible to examine the relationship between the lengths of his narrow rhythm units (in syllables) and the mean lengths of syllables (in phonemes), and also to examine whether the lengths of Jassem's units follow a regular probability distribution.

In relation to the second possibility mentioned above, that both annotators failed to delimit the intonation units correctly in many of their texts, it stretches credulity to believe that two trained and experienced phoneticians made so many errors in locating intonation unit boundaries that it affected the fitting of probability distributions to half or more of their texts. Certainly, Knowles (1991) uncovered quantitative differences between himself and BJW in their recognition of intonation units, with Knowles recognizing a greater number of intonation units than BJW in the overlap texts of the MARSEC corpus. He attributed this to a greater tendency on his part to respond to "minor boundary phenomena", such as pitch movements. Bearing in mind that pitch movements, in the guise of nuclear tones, are the defining feature of intonation units in most versions of the 'British' approach, this may account for the slightly greater success in modelling his analyses than those of BJW. However, it seems unlikely that GOK's annotations would be so inconsistent as to cause problems with the other half or more of his texts when the Altmann-Fitter and Menzerath-Altmann equation were applied to them.

A final possibility raised above was that the intonation unit may not, in fact, be a legitimate language construct at all within a constituency hierarchy of prosodic units. This experiment alone cannot answer decisively whether this statement is true or false; it can merely say that there is very little evidence here for a constituency relationship between Abercrombie's (1964) feet and the higher level of intonation units, as operationalized by either of the two Aix-MARSEC annotators. However, the findings of the experiment would certainly be consistent with the truth of the statement. It is worth noting in this regard that Knowles (1991, p. 160) concluded by seeking to de-couple prosodic boundaries from notions of constituency: "Discontinuities must ... be identified and labeled in their own account, and not bound to the kind of tone group theory that underlies conventional transcription".

To sum up, this experiment has cast further doubt on the hypothesis that syllables, feet, and intonation units are contiguous levels of a prosodic constituency hierarchy. In contrast to other linguistic constituency phenomena (e.g. in syntax), it has not been possible to establish a lawful relationship between the lengths of intonation units in feet and the mean lengths of the constituent feet in syllables. Equally, it has not been possible to find a regular probability distribution for the frequencies of length classes of intonation units measured in feet, whereas this has so far been universally possible for other linguistic units when measured in terms of their proper constituents. It remains an open question as to whether the problem is merely with the constituency hierarchy (for instance, whether additional levels of structure exist between the foot and the intonation unit) or whether the picture is complicated by problems in the definition and delimitation of units, which leads to chaotic results when mathematical modelling is applied. This latter problem could be approached by applying the same quantitative linguistics methodology to other prosodically annotated spoken texts. In this connection, it would be of value to check and, where necessary, enhance the annotation quality of prosodic boundaries so that they match the formal 'British school' definition as closely as possible - cf. Jassem's (1996) analysis of nuclear tones.

Future work could also usefully examine other prosodic hierarchies using the quantitative linguistics framework outlined here - e.g., the hierarchical constituency model outlined by Ladd (1986) - although

this might entail considerable work in annotating transcribed spoken data with new categories of structure.

## **Acknowledgements**

I am grateful to Gabriel Altmann (Bochum/Lüdenscheid) for his comments on an early draft of this paper.

The data used in this study is part of the Aix-MARSEC database, version 2, courtesy of the Speech & Language Data Repository (SLDR), <http://sldr.org>, identifier: sldr000033.

## References

Abercrombie, D. 1964. Syllable quantity and enclitics in English. In Abercrombie, D., Fry, P. MacCarthy, N., Trim, J. (Eds.), *In Honour of Daniel Jones*, Longman, London, pp. 216-222.

Altmann-Fitter. 1997. *Iterative Fitting of Probability Distributions*. RAM-Verlag, Lüdenscheid.

Altmann, G. 1980. Prolegomena to Menzerath's Law. In Grotjahn, R. (Ed.), *Glottometrika 2*, Brockmeyer, Bochum, pp. 1-10.

Altmann, G. 1992. Das Problem der Datenhomogenität. In: Rieger, B. (Ed.), *Glottometrika 13*, Brockmeyer, Bochum, pp. 287-298.

Altmann, G. 2008. On the symbiosis of physicists and linguists. *Romanian Rep. Phys.* 60, 417-422.

Altmann, G., Köhler, R. 2005. Aims and methods of quantitative linguistics. In Altmann, G., Levickij, V., Perebyinis, V. (Eds.), *Problems of Quantitative Linguistics*, Ruta, Černivci, pp. 12-41

Altmann, G. & Meyer, P. 2005. Physicist's look at language. In Altmann, G., Levickij, V., Perebyinis, V. (Eds.), *Problems of Quantitative Linguistics*, Ruta, Černivci, pp. 42-59.

Altmann, G., Schwibbe, M.H. 1989. *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Olms, Hildesheim.

Andres, J. 2010. On a conjecture about the fractal structure of language. *J. Quant. Ling.* 17, 101-122.

Andres, J., Benešová, M., Kubáček, L., Vrbková, J. (2012). Methodological note on the fractal analysis of

texts. *J. Quant. Ling.* 19, 1-31.

Antić, G., Stadlober, E., Grzybek, P., Kelih, E. 2006. Word length and frequency distributions in different text genres. In Spiliopoulou, M., Kruse, R., Nürnberger, A., Borgelt, C., Gaul W. (Eds.), *From Data and Information Analysis to Knowledge Engineering*, Springer, Heidelberg, pp. 310-317.

Auran, C., Bouzon, C., Hirst, D.J. 2004. The Aix-MARSEC project: an evolutive database of spoken English. In Bel, B., Marlien, I. (Eds.), *Proceedings of the Second International Conference on Speech Prosody*, Nara, Japan, pp. 561-564.

Barney, T. 1999. Readers as text processors and performers: A new formula for poetic intonation. *Discourse Process.* 28, 155-167.

Best, K.-H. (Ed.). 2001. *Häufigkeitsverteilungen in Texten*. Peust & Gutschmidt, Göttingen.

Best, K.-H. 2002. The distribution of rhythmic units in German short prose. *Glottometrics* 3, 136-142.

Best, K.-H. 2005. Längen rhythmischer Einheiten. In Köhler, R., Altmann, G., Piotrowski, R. (Eds.), *Quantitative Linguistik - Quantitative Linguistics. Ein internationales Handbuch*. Mouton de Gruyter, Berlin, pp. 208-214.

Best, K.-H. 2011. Bibliographische Übersicht zum Göttinger Projekt zur Quantitativen Linguistik. Retrieved July 25, 2011, from <http://wwwuser.gwdg.de/~kbest/litlist.htm>

Bouzon, C., Hirst, D. 2004. Isochrony and prosodic structure in British English. In Bel, B., Marlien, I. (Eds.), *Proceedings of the Second International Conference on Speech Prosody*, Nara, Japan, pp. 223-226.

- Brierley, C., Atwell, E. 2011. Non-traditional prosodic features for automated phrase break prediction. *Lit. Ling. Comput.* 26, 279-284.
- Bunge, M. 1977. *Ontology*. Reidel, Dordrecht.
- Čech, R., Popescu, I.-I., Altmann, G. 2011. Word length in Slovak poetry. *Glottometrics* 22, 44-56.
- Clayman, D.L 1981. Sentence length in Greek hexameter poetry. In Grotjahn, R. (Ed.), *Hexameter Studies*, Brockmeyer, Bochum, pp. 107-136.
- Cramer, I.M. 2005. The parameters of the Altmann-Menzerath Law. *J. Quant. Ling.* 12, 41-52.
- Cruttenden, A. 1969. Review of Halliday 1967. *J. Ling.* 5, 309-315.
- Crystal, D. 1969. Review of Halliday 1967. *Language* 45, 378-393.
- Grzybek, P. 2006. History and methodology of word length studies: The state of the art. In Grzybek, P. (Ed.), *Contributions to the Science of Text and Language: Word Length Studies and Related Issues*, Springer, Dordrecht, pp. 15-90.
- Halliday, M.A.K. 1967. *Intonation and Grammar in British English*. Mouton, The Hague.
- Halliday, M.A.K., Greaves, W.S. 2008. *Intonation in the Grammar of English*. Equinox, London.
- Ihaka, R., Gentleman, R. 1996. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299-314.

Jassem, W. 1952. Intonation of conversational English (educated Southern British). Nakładem Wrocławskiego Towarzystwa Naukowego, Wrocław.

Jassem, W. 1996. A quantitative analysis of standard British English nuclear tones. *J. Quant. Ling.* 3, 229-243.

Kelih, E. 2006. Häufigkeiten von Graphemen und Lauten: Zwei Ebenen - ein Modell? (Re-Analyse einer Untersuchung von A.M. Peškovskij). In Grzybek, P., Köhler, R. (Eds.), *Exact Methods in the Study of Language and Text*, Mouton de Gruyter, Berlin, pp. 267-277.

Knight, R. 2013. Laws governing rank frequency and stratification in English texts. *Glottometrics* 25, 30-42.

Knowles, G. 1991. Prosodic labelling: the problem of tone group boundaries. In Johansson, S., Stenstrom, A.-B. (Eds.), *English Computer Corpora: Selected Papers and Research Guide*, Mouton de Gruyter, Berlin, pp. 149-163.

Knowles, G. 1992. Pitch contours and tones in the Lancaster/IBM Spoken English Corpus. In Leitner, G. (Ed.), *New Dimensions in Corpus Linguistics*, Mouton de Gruyter, Berlin, pp. 289-299.

Knowles, G. 1994. Annotating large speech corpora: Building on the experience of Marsec. *Hermes J. Ling.* 13, 87-98.

Knowles, G., Wichmann, A., Alderson, P. (Eds.). 1996. *Working with Speech: Perspectives on Research into the Lancaster/IBM Spoken English Corpus*. Longman, London.

Köhler, R., Altmann, G, Piotrowski, R. (Eds.) 2005. *Quantitative Linguistik - Quantitative Linguistics*. Ein



internationales Handbuch. Mouton de Gruyter, Berlin.

Köhler, R. 1986. Zur linguistischen Synergetik: Struktur und Dynamik der Lexik. Brockmeyer, Bochum.

Köhler, R. 1987. Systems theoretical linguistics. *Theor. Linguist.* 14, 241-257.

Köhler, R., Altmann, G. 1997. On explorative methodology concerning probability distributions. In *SoftStat'97: Advances in Statistical Software* 6, pp. 335-339.

Ladd, D.R. 1986. Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook* 3, 311-340 .

Marbe, K. 1904. Über den Rhythmus der Prosa. J. Ricker'sche Verlagsbuchhandlung, Giessen.

Menzerath, P. 1954. Die Architektonik des deutschen Wortschatzes. Dümmler, Bonn.

Popescu, I.-I., Altmann, G., Grzybek, P., Jayaram, B.D., Köhler, R., Krupa, V., Mačutek J., Pustet, R., Uhlířová, L., Vidya, M.N. 2009. *Word Frequency Studies*. Mouton de Gruyter, Berlin.

Popescu, I.-I., Mačutek, J., Altmann, G. 2009. *Aspects of Word Frequencies*. RAM-Verlag, Lüdenscheid.

Rietveld, T., van Hout, R., Ernestus, M. 2004. Pitfalls in corpus research. *Comput. Humanities* 38, 343-362.

Roach, P., Knowles, G., Varadi, T., Arnfield, S. 1993. MARSEC: A machine-readable spoken English corpus. *J. Int. Phon. Assoc.* 23, 47-54.

Taylor, L.J., Knowles, G. 1988. Manual of Information to Accompany the SEC Corpus: The Machine-Readable Corpus of Spoken English. Retrieved July 2, 2013, from <http://icame.uib.no/sec/>

Wimmer, G., Altmann, G. 1999. Thesaurus of Univariate Discrete Probability Distributions. Stamm, Essen.

Ziegler, A., Altmann, G. 2003. Text stratification. *J. Quant. Ling.* 10, 275-292.