

# MEASURING THE EUPHONY

Petr Plecháč – Jakub Říha

In our paper we would like to present the first results of applying the algorithm for the automatic analysis of euphony developed by Gabriel Altmann (1966a). By euphony we mean, like Altmann, the aesthetically relevant repetition of sounds in line. On the one hand, we expand the extent of this term – as we do not utilize the usual differentiation between euphony and cacophony – and narrow it down on the other hand, for our definition does not include repetitions of groups of sounds, structures that are superior to the line, etc. As far as we know, Altmann's algorithm has only been applied to small sets of texts (ALTMANN 1966b, ČECH–POPESCU–ALTMANN 2011). Thus, our experiment most likely represents the first attempt at its application for analyzing an extensive corpus. (We have applied the algorithm when analyzing over 80,000 poems that contained over 2,000,000 lines.)

Euphony in general is based on the deviation in the distribution of certain sounds from the extent of language probability. For this reason, Altmann proposes the following procedure as one of the possible ways of quantifying euphony (or one of its possible manifestations). The algorithm is based on the known frequency of individual sounds in the language and proceeds with each individual line. Based on its frequency, the probability of its repetition – or the probability that the given sound will occur  $x$ -times or more times is computed for each repetition in each line.

$$(1) \quad P(X \geq x_i) = \sum_{x=x_i}^N \binom{N}{x} p^x (1-p)^{N-x}$$

Probability that vowel/consonant will occupy  $x$  and more positions out of  $N$  vocalic/consonantic positions  
( $p$  = probability of occurrence of vowel/consonant).

If the probability is  $< 0.05$  (i.e. the conventional significance level  $\alpha$ ), the given repetition of the sound is considered to be euphonicly relevant and is assigned a euphonic coefficient  $\varepsilon$  based on the subtraction of these two values.

$$(2) \quad \varepsilon = \begin{cases} 100[\alpha - P(X \geq x_i)] & \text{if } \alpha > P(X \geq x) \\ 0 & \text{otherwise} \end{cases}$$

The euphonic coefficient of the entire line ( $e$ ) is computed as the mean value of euphonic coefficients of all relevant repetitions.

$$(3a) \quad e = \begin{cases} \frac{1}{k} \sum_{i=1}^k \varepsilon_i & \text{if } k \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

The euphonic coefficient of the entire poem ( $E$ ) is computed as the mean value of euphonic coefficients of individual lines.

$$(4) \quad E = \frac{1}{n} \sum_{j=1}^n e_j$$

We have slightly modified Altmann's procedure for our needs: the euphonic coefficient of the line was not computed as the *mean value* of coefficients of relevant repetitions but as their *sum total*.

$$(3b) \quad e = \sum_{i=1}^k \varepsilon_i$$

In our opinion, the final value is inappropriately affected by marginal configurations that may result from e.g. parallelism when applying the first above-mentioned procedure.

Let us compare the last two lines from the poem *Kostelní hlahol zval horaly* by Adolf Racek (EXAMPLE 1), in which a relevant repetition of consonants [b] and [l] is found with almost identical probability of occurrence. In the last line, moreover, the repetition of the long vowel [i:] occurs. This vowel, which obtains only a low euphonic coefficient in Czech due to its relatively high frequency, decreases the total coefficient of the line in Altmann's concept. As a consequence, its value is lower than the value of the previous line by more than one third. Yet, the value of the euphonic coefficient of the two lines would be virtually identical if this vowel had not been present.

Our experiment has yielded considerably satisfying results. As expected, the euphonic coefficient obtained the highest values primarily in symbolist poems and poems written by authors who had been influenced by symbolism. Partial tests showed that the algorithm was capable of detecting relevant sound structures.

Let us present the above-mentioned poem *Kostelní hlahol zval horaly* by Adolf Racek (EXAMPLE 1) and *Zvony* by František Leubner (EXAMPLE 2) as examples of poems that have obtained the highest values. We can see that the poets used different methods to achieve euphony. While the total value of the euphonic coefficient in the poem by Racek is constituted by a consonant [l] by almost 50%, where other sounds serve only as accessories to this euphonic frame, in the poem by Leubner the final coefficient composition is very heterogeneous with no dominant sound or sounds. Thus, euphony is a function of the entire text in the former case and a function of the individual lines in the latter one. The variability of data thus could serve as one of the starting points for outlining the basic euphonic typology in the future.

Allow us to add that the experiment has also detected partial weak points in the algorithm. First of all, the repetition of units on higher linguistic levels is not taken into account when marking sound repetitions. Thus, the quatrain by Josef Svatopluk Machar (EXAMPLE 3) has been classified among the texts with the highest euphonic coefficient. However, one would be reluctant to mark it as euphonicly relevant. The high value is caused primarily by several repetitions of the word *guma* (*rubber*), which contains one of the least frequented consonants [g]. (Unlike Russian, Czech does not have the original proto-Slavic [g]. The [g] > [h] shift took place as early as the 13<sup>th</sup> century. Thus, [g] nowadays occurs only in loanwords.) For this reason, we carried out the experiment for the second time, with a slight adjustment: the program takes note of only one occurrence in cases when a full word (or its forms) occurs more than once in a line. For instance, when analyzing the above-mentioned lines by Machar:

Duch je guma, páteř guma, guma přesvědčení  
guma prospěch republiky, nad gumu dnes není

the first occurrence of the word *guma* in each line is observed. No relevant euphonic structure has been found:

Duch je guma, páteř [...] [...] přesvědčení  
guma prospěch republiky, nad [...] dnes není.

The parameters that have been set up in this way have pushed many similar (irrelevant) cases away from the top ranking. However, one can still find texts the euphonic value of which can be considered disputable at least among poems with a high euphonic coefficient. In such texts, repetition of sounds is not caused by the repetition of identical words but by the repetition of a word and its derivatives. For instance, the final euphonic coefficient in the poem *Fragment z pozůstalosti* by Stanislav Kostka Neumann is caused to a large degree by the repetition of lines in which the words *rodič* (*parent*) and *prarodič* (*grandparent*) occur:

Moji rodiče a prarodiče byli Černoši...  
Moji rodiče a prarodiče byli Indiáni...

Unfortunately, we are not currently able to detect word-forming relations automatically between the individual words. A satisfactory solution for such situations still remains to be found.

Our third – and last – step focused on automatic detection of cases of the so-called sound irradiation, i.e. a situation when the sounds included in the designation of the central motif or in another key word serve as chief euphony carriers. For this reason, we modified the algorithm in the following way: first of all, the most frequently repeated word was detected in each poem (the minimum determined as three occurrences; only one occurrence in the line was taken note of for the above-mentioned reasons). Attention was paid only to consonants that occurred in some form of this word. Vowels were not taken into consideration, for the set of all forms of a single word mostly contains the entire list of Czech vowels due to the developed inflection and frequent alternations in the word base. From now on the euphonic coefficient assessed for such consonants will be called *irradiation coefficient*.

When analyzing irradiation, one naturally faces the same problems as when analyzing euphony. A high irradiation coefficient has been assigned e.g. to the above-mentioned *Fragment* by Neumann with *rodič* as the key word and all occurrences of the word *prarodič* assigned as its intense irradiation. Despite all these drawbacks, the algorithm detected many relevant cases.

Allow us to name the poem *Já nejsem smuten* by Jaroslav Kolman Cassius (EXAMPLE 4). The most frequented word is *smutný* (*sad*); it is repeated seven times in various forms and can be considered the central motif of the entire poem. At the same time inherent consonants, i.e. [s][m][t][n] forms noticeable euphonic structures in the poem.

As we have seen, this approach does not lack errors and drawbacks. Apart from morphemic composition, other factors should be taken into consideration as well, e.g. euphony that occurs only in a part of the text, the repetition of entire sound groups, sound structures that are based on the alternation of strong and weak positions of the meter, sound structures that are superior to the line, etc. Some procedures that reflect many of the above-mentioned cases have already been devised (see WIMMER et al.: 55–85). We

believe that the probability analysis presented herein could yield precious results in the future. Precise euphony quantification should enable us to avoid the immense subjectivity that usually accompanies the research, as well as compare and classify the obtained data – either on the level of the individual authors, poetic schools, generations, or even entire national versifications.

ALTMANN, G.

1966a “The Measurement of Euphony”, in: *Teorie verše I* (Brno: UJEP), s. 259–261

1966b “Binomial Index of Euphony for Indonesian Poetry”, *Asian and African Studies* 2, s. 62–67

ČECH, R. – POPESCU, I. I. – ALTMANN, G.

2011 “Euphony in Slovak lyric poetry”, *Glottometrics* 22, s. 5–16

WIMMER, G. – ALTMANN, G. – HŘEBÍČEK, L – ONDREJOVIČ, S. – WIMMEROVÁ, S.

2003 *Úvod do analýzy textov* (Bratislava: VEDA)