Petr Plecháč, Jakub Říha
(Praha, Ústav pro českou literaturu AV ČR)

Measuring euphony

This paper presents initial results of applying the algorithm for the automatic analysis of euphony developed by Gabriel Altmann (Altmann 1966b; 1966a). By euphony we mean, following Altmann, the aesthetically relevant repetitions of sounds in line. On the one hand, we expand the scope of this term — as we do not utilize the usual differentiation between euphony and cacophony; on the other hand, we narrow it down, 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 number of texts (Ibid.; Čech 2001). Thus, our experiment most likely represents the first attempt at its application for analyzing an extensive poetic 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 possible ways of quantifying euphony (or one of its possible manifestations). The algorithm is based on 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 (Formula 1) — is computed for each repetition in each line.

Probability that vowel/consonant occupies *x* and more positions out of *N* positions (*p* = probability of occurence of vowel/consonant)

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

If the probability is — 0.05 (i.e. the conventional significance level), the given repetition of the sound is considered to be euphonically relevant and is assigned a euphonic coefficient ε based on the subtraction of these two values (Formula 2).

Euphonic coefficient of a vowel / consonant in a particular line (*a* = 0,05)

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

The euphonic coefficient of the entire line (e) is computed as the mean value of euphonic coefficients of all relevant repetitions (Formula 3a); the euphonic coefficient of the entire poem (E) is computed as the mean value of euphonic coefficients of individual lines (Formula 4).

Euphonic coefficient of a line

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

Euphonic coefficient of a poem

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

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 (Formula 3b).

Euphonic coefficient of a line

(FORMULA 3b) $$e=\sum_{i=1}^{k}\varepsilon_i$$ [PLECHÁČ–ŘÍHA]

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

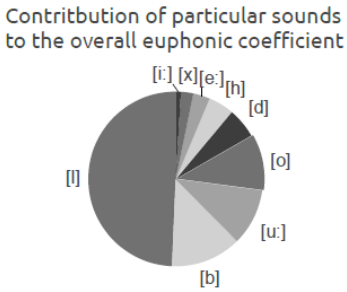
Let us compare the last two lines from the poem Kostelni hlahol zval horaly by Adolf Racek (Example1), in which a relevant repetition of consonants [b] and [l] is found with almost identical probability of occurrence. In the last line, moreover, the repetition of the long vowel [i:] occurs.

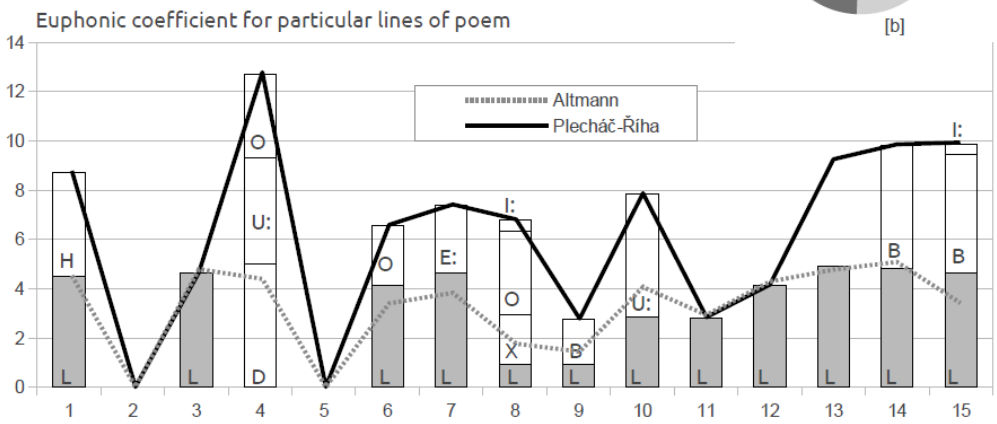
(EXAMPLE 1)
KOSTELNÍ HLAHOL ZVAL HORALY ***E*** **= 6,214**
(Adolf Racek, *U základů*, 1914)

1 Kostelní hlahol zval horaly	kostelni: hlahol zval horali
2 k oblakům, zrudlé jež doutnaly,	koblaku:m zrudle: jeʃ doʊ̯tnali
3 unylé flétny volaly;	unile: fle:tni volali
4 do údolí, do údolí.	dou:doli: dou:doli:
5 U splavu blouznivě topoly	usplavu bloʊ̯znivje topoli
6 rozkolébaly kol vrcholy,	roskole:bali kol vrxoli
7 kouzelné flétny dál hlaholí	koʊ̯zelne: fle:tni dal hlaholi:
8 ospalých do lích, do polí	ospali:x doli:x dopoli:
9 bláhovém mládí o bolu,	bla:hove:m mla:ji: obolu
10 o sladkém úlů úkolu,	oslatke:m u:lu u:kolu:
11 stmělý dol až sklenné bílo	stmɲeli: dol aʃ sklene: bi:lo
12 luny zahalilo.	luni zahalilo
13 Alt flétny sola do zmlklých rolí	alt fle:tni sola dozmlkli:x roli:
14 blábolí; mládí nebylé bolí,	bla:boli: mla: ji: nebile: boli:
15 bolí, bolí, láskou bolí.	boli: boli: la:skoʊ̯ boli:

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 had this vowel not been presented.





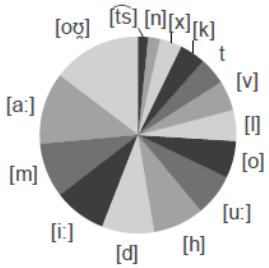
Our experiment has yielded considerably satisfying results. As expected, the euphonic coefficient obtained the highest values primarily in symbolist poems and poems written by authors who had been influenced by symbolism. Partial tests showed that the algorithm was capable to detect relevant sound structures.

Let us present the above-mentioned poem Kostelni 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.

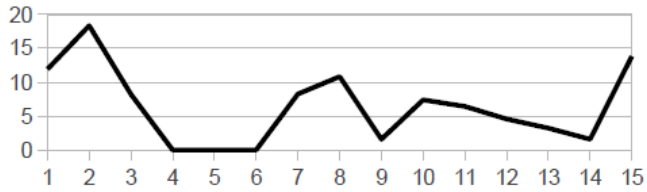
(EXAMPLE 2)
ZVONY *E* = 6,408
(František Leubner, *Na okrajích kancionálu života*, 1897)

1	Vím vám dum doby: Dumou mou duní nad Prahou zvony,	vi:m va:m dům doby dumoʊ̯ moʊ̯ duɲi: natprahoʊ̯ zvoni
2	a kvílí – kvílí – kvílí – Ví-li, co kvílí? –	a kvi:li: kvi:li: kvi:li: vi:li tso kvi:li:
3	kdy v hloubi touha zdlouha napotom chvílí	kdi v hloʊ̯bi toʊ̯ha zdloʊ̯ha napotom xvi:li
4	v nich údery bolu lká na hony stony...	vɲix u:deri bolu lka: nahoni stoni
5	Dóm, – rudý kol západ, – hrad, – mlhy clony,...	do:m rudi: kol za:pat hrat mlhi t͡sloni
6	přes jezy ječí Vltava, v temnou dálku pílí,	pr̥es jezi jet͡ʃi: vltava ftemnoʊ̯ da:lku pi:li:
7	a onde ostrovů dolů sklony	a onde ostrovu: dolu: skloni
8	se do vody chýlí, kam dům, dům se zrcadlí bílý.	se dovodi xi:li: kam du:m du:m se zr̥t͡sadli: bi:li:
9	Od Sionu shodou tonů tonu v rekviem hran...	odsijonu sxodoʊ̯ tonu: tonu v rekvijem hran
10	Jich pádný hlahol neumlká, lká a lká v chladný vzduch	jix pa:dni: hlahol neumlka: lka: a lka: fxladni: vzdux
11	a stíny honí od zvonic jak havranů hejno, hejno vran.	a sci:ni honi: odzvoɲit͡s jak havranu: hejno hejno vran
12	Pak klekání vlání ustává pěti, –	pak kleka:ɲi: vla:ɲi: usta:va: pjeci
13	snad budí je ráno sám Pán Bůh...	snat buji: je ra:no sa:m pa:n bux
14	Noc orlice černá, od západu letí	not͡s orli t͡se tʃerna: otza:padu leci:
15	tmou – tmou – tmou – –	tmoʊ̯ tmoʊ̯ tmoʊ̯

Contritbution of particular sounds
to the overall euphonic coefficient



Euphonic coefficient for particular lines of poem



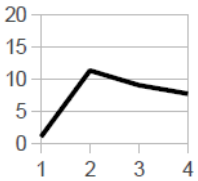
We can see that poets used different methods to achieve euphony. While the total value of the euphonic coefficient in the poem by Racek is constituted by a consonant [l] by almost 50%, where other sounds serve only as accessories to this euphonic frame, in the poem by Leubner the final coefficient composition 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 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.

Besides, 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.

GUMA ***E*** **= 7,312**
(Josef Svatopluk Machar, *Na okraj dnů*, 1935)

1 Na sever, na východ, na západ, v jih –	nasever navi:xot naza:pat vjix
2 agrárník – klerikál – pokrokář v mih –	agra:rɲi:k klerika:t pokroka:r vmix
3 duch je guma, páteř guma, guma přesvědčení,	dux je guma pa:teř guma guma přesvjet͡ʃeɲi
4 guma prospěch republiky, nad gumu dnes není!	guma prospjex republiki nadgumu dnes neɲi:

Euphonic coefficient for particular lines of poem



However, one would be reluctant to mark it as euphonically 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 13th 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 eliminated many similar (irrelevant) cases 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 individual words. A satisfactory solution for such situations still remains to be found.

Our third — and the last — step focused on automatic detection of cases of the so-called sound irradiation, that is, 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, the most frequently repeated word was detected in each poem (the minimum determined as three occurrences; only one occurrence in the line was counted 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, for example, 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.

Let us proceed to the poem Ja nejsem smuten by Jaroslav Kolman Cassius (Example 4).

The most frequented word is smutny (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 [s][m][t][n] form noticeable euphonic structures in the poem.

(EXAMPLE 4)
JÁ NEJSEM SMUTEN
(Jaroslav Kolman Cassius, *Ovčín*, 1937)

Já nejsem smuten, to jen písně moje, zaklínající stesk mých opožděných žní. Pozdního zrání zpěv, šum zbloudilého roje i píseň podzimní melancholicky zní.	ja: nejsem smuten to jen pi:sɲe moje zakli͡naji:t͡si: ste͡sk mi:x opoʒd̪eni:x ʒɲi: pozdɲi:ho zra:ɲi: spjev ʃum zbloʊ̯ile:ho roje i pi:seɲ podzimɲi: melanxolicki zɲi:
Já nejsem smuten, to jen písně moje melodii snad mají bolavou, když protrhnou se v srdci staré zdroje a zatopí je jarní záplavou.	ja: nejsem smuten to jen pi:sɲe moje melodiji snat maji: bolavoʊ̯ kdiʃ prot̪rhnoʊ̯ se v s̪t͡si stare: zdroje a zatopi: je jarɲi: za:plavoʊ̯
Já nejsem smuten. Nemám žádný žal. Má píseň radostně jak jarní míza tryská. Jak smutná byla krása, kráčící opodál, oč smutnější je blízká, tak strašidelně blízká!	ja: nejsem smuten nemá:m ʒa:dni: ʒal ma: pi:seɲ radost̪ne jak jarɲi: mi:za t̪ri͡ska: jak smutna: bila kra:sa kra:t̪ʃi:t͡si: opoda:l oʧ smut̪nejʃi: je bli:ska: tak straʃidelɲe bli:ska:
V jezero azurové, radosti, vnoř sítě a smutnou rybku ulovíš ze hlubin. Obejmi plamen a popel políbí tě. Hory a oblaka i křídla mají stín.	v jezero azurove: radosci vnoɾ̪ si:ce a smutnoʊ̯ ripku uloviːʃ zehlubi͡n obejmi plamen a popel poli:bi: ce hori a oblaka i kɾi:dla mají: sci:n
Radostně volám v slunný krásy kraj své štěstí jako prozpěvují ptáci. Já nejsem smuten, to jen stinný háj mou píseň tesknou ozvěnou mi vrací.	radost̪ne vola:m v sluni: kra:si kraj sve: ʃcesci: jako prospjevuji: pta:t͡si ja: nejsem smuten to jen sci:ni: ha:j moʊ̯ pi:seɲ teskn̪oʊ̯ ozvjenoʊ̯ mi vraci:

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. For example, 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 (Wimmer 2003, p. 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 significant element of subjective evaluation 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.

Bibliography

Altmann 1966a: Altmann, G. Binomial Index of Euphony for Indonesian Poetry // Asian and African Studies. 2. — Bratislava, 1966. — S. 62–67.

Altmann 1966b: Altmann, G. The Measurement of Euphony // Teorie verše. I. — Brno, 1966. — S. 259–261.

Čech 2001: Čech, R.; Popescu, I.I.; Altmann, G. Euphony in Slovak lyric poetry // Glottometrics. — 2001. — №22. — S. 5–16.

Wimmer 2003: Wimmer, G., Altmann, G., Hřebíček, L., Ondrejovič, S., Wimmerová, S. Úvod do analýzy textov. — Bratislava, 2003.