

Tra-la-Lyrics 2.0: Automatic Generation of Song Lyrics on a Semantic Domain

Hugo Gonçalo Oliveira

*CISUC, Department of Informatics Engineering
University of Coimbra
Portugal*

HROLIV@DEI.UC.PT

Editor: Tarek R. Besold, Kai-Uwe Kühnberger, Tony Veale

Abstract

Tra-la-Lyrics is a system that generates song lyrics automatically. In its original version, the main focus was to produce text where stresses matched the rhythm of given melodies. There were no concerns on whether the text made sense or if the selected words shared some kind of semantic association. In this article, we describe the development of a new version of Tra-la-Lyrics, where text is generated on a semantic domain, defined by one or more seed words. This effort involved the integration of the original rhythm module of Tra-la-Lyrics in PoeTryMe, a generic platform that generates poetry with semantically coherent sentences. To measure our progress, the rhythm, the rhymes, and the semantic coherence in lyrics produced by the original Tra-la-Lyrics were analysed and compared with lyrics produced by the new instantiation of this system, dubbed Tra-la-Lyrics 2.0. The analysis showed that, in the lyrics by the new system, words have higher semantic association among them and with the given seeds, while the rhythm is still matched and rhymes are present. The previous analysis was complemented with a crowdsourced evaluation, where contributors answered a survey about relevant features of lyrics produced by the previous and the current versions of Tra-la-Lyrics. Though tight, the survey results confirmed the improvements of the lyrics by Tra-la-Lyrics 2.0.

Keywords: computational creativity, linguistic creativity, song lyrics, poetry, rhythm, semantics

1. Introduction

Creative systems exhibit behaviours that would be deemed as creative by unbiased observers (Colton and Wiggins, 2012), often rendered in the form of visual art, music, or pieces of text, among others. Computational attempts to linguistic creativity include the generation of verbally-expressed humor (Binsted and Ritchie, 1994), narratives (Gervás et al., 2005), metaphors (Veale and Hao, 2008), neologisms (Smith, Hintze, and Ventura, 2014), slogans (Tomašič, Žnidaršič, and Papa, 2014) or poetry (Gervás, 2001). Though also explored, the generation of song lyrics has received less attention. In addition to the textual component, the previous task involves a musical aspect.

The generation of poetry and song lyrics aims at producing natural language text with a creative goal, either for entertainment purposes, for studying the process of composition from a computational point of view, or both. In fact, song lyrics are a kind of poetry that aims to be sung, instead of just read. Therefore, the lyric's text is more closely attached to the song's rhythm, while in classic forms of poetry text is attached to the form (e.g. a sonnet). In the same lyrics, it is common to have lines of different lengths, grouped in stanzas of different sizes. On the other hand, the lines of a classic poem tend to have a regular metre (same number of syllables) as well as some regularity

in the occurrence of rhymes and lines per stanza. Of course, this does not account for alternative kinds of poetry with free verse, or with a shape.

Tra-la-Lyrics (Gonçalo Oliveira, Cardoso, and Pereira, 2007b,a) was developed to produce text based on a given rhythm, and was applied to the generation of song lyrics in Portuguese. Although when using only rhythmic constraints, the produced text frequently matches the rhythm, this is achieved without any considerations on semantics. On the other hand, PoeTryMe (Gonçalo Oliveira, 2012; Gonçalo Oliveira and Cardoso, 2015) is a generic platform for poetry generation that produces text without any explicit notion of rhythm, but where semantically coherent sentences are organised according to predefined poetry forms. PoeTryMe has a high level of customisation and its constituents can be adapted towards the creation of significantly different poetry generation systems.

In order to generate song lyrics with semantic considerations, we took advantage of the modular architecture of PoeTryMe and integrated the main component of the original Tra-la-Lyrics there. The reported work can be seen as an upgrade to PoeTryMe, which, in addition to previous features, can now assess the rhythm of generated text. However, since the new instantiation has the same specific goal as Tra-la-Lyrics – to generate lyrics with a given rhythm – and reuses its main component, it was baptised as Tra-la-Lyrics 2.0. The main difference is that lyrics are now produced on a semantic domain, drawn from a set of seed words.

To analyse the progress of Tra-la-Lyrics 2.0 towards its previous versions, we focused on three features – rhythm, rhymes and semantics – and compared their occurrence in lyrics by the new system and by two generation strategies of the original Tra-la-Lyrics. The main conclusions were that, although the rhythm matches are not as high as in the most simple strategy of Tra-la-Lyrics and rhymes are also less, both of these features are still clearly present. On the other hand, based on word association measures, we can say that Tra-la-Lyrics 2.0's text is more semantically associated to the initial seed words, which was one of our main goals. Moreover, though with tight differences, a crowdsourced survey confirmed the improvements of the lyrics produced by Tra-la-Lyrics 2.0, not only concerning semantics but, more surprisingly, also rhythm and sound.

This article starts by enumerating related work on computational approaches to the generation of poetry and song lyrics. After that, the original systems used, Tra-la-Lyrics and PoeTryMe, are described. Detailed information on the integration of the previous systems is then provided, together with illustrative examples of generated lyrics. Before concluding, a comparison of the rhythm, rhymes and semantic similarity in lyrics produced by different strategies of Tra-la-Lyrics is discussed, and followed by the results of a survey, answered by several contributors.

2. Related Work

The automatic generation of poetry became a research topic in the scope of artificial intelligence after earlier works of Gervás (2001) and Manurung (2003). In poetic text, features like metre, rhyme, alliteration and figurative language play an important role, which sets the main challenge on coping with different levels of language. Though, not all of them must be strictly addressed. For instance, syntactic or semantic rules may sometimes be relaxed to satisfy metre constraints. Since the first poetry generation systems, a variety of paradigms has been applied to this task, including case-based reasoning (Gervás, 2001), evolutionary algorithms (Manurung, 2003), constraint programming (Toivanen, Jarvisalo, and Toivonen, 2013), or multi-agent systems (Miszta and Indurkha, 2014).

Although at different levels, most poetry generation systems use a template representation of the poems and their lines, typically acquired from human-created poems, where some fragments can be replaced, thus originating a broader variety of constructions. When semantic constraints are present, and they often are, the system tries to replace those parts by other semantically coherent text fragments, while satisfying the metre and other poetic features. In several systems, generation starts with a prose message (Gervás, 2001), a theme (Toivanen, Järvisalo, and Toivonen, 2013) or a set of seed words (Netzer et al., 2009), which constrain the poem search space and set the semantic domain. The choice of relevant words/expressions may be achieved either with the help of semantic knowledge bases (Agirrezabal et al., 2013), by exploring models of word associations extracted from corpora (Netzer et al., 2009; Toivanen, Järvisalo, and Toivonen, 2013), or both (Colton, Goodwin, and Veale, 2012). However, with few exceptions, the previous approaches ignore the rhythm of the lines, or deal with it in a simplified way. One of the few exceptions that consider semantics, syntax and rhythm in poetry is Chrismartin, Tobing, and Manurung (2015). In addition to the rhythmic constraints, they relied on syntactic predicates to generate meaningful poetry and exploited a semantic knowledge base to retrieve synonyms and increase the ability to paraphrase. However, the authors admit that considering so many features results in too much complexity and a long time for producing a poem.

After the first approaches to the automatic generation of poetry, some attention has been given to the automatic generation of song lyrics. Tra-la-Lyrics (Gonçalo Oliveira, Cardoso, and Pereira, 2007b,a) was presented as a system for generating text according to the rhythm of a melody. It tries to match the stress of the words with the strong beats of the melody and, although a rough meaning might sometimes emerge from the lyrics, none of the implemented generation strategies dealt with any kind of semantics. More on Tra-la-Lyrics is presented in section 3.1.

Besides rhythm and rhymes, and similarly to poetry generation approaches, other attempts to generate song lyrics considered also syntax, in the form of valid part-of-speech (POS) sequences, and semantics, by restricting the selection of words by some relatedness measure. Ramakrishnan A and Devi (2010) generate song lyrics in two steps: (i) drawing possible syllable patterns for the melody; (ii) with the help of a semantic knowledge base, generate meaningful sentences that match those patterns. Barbieri et al. (2012) learn a constraint model from a set of lyrics of the same author and then generate new lyrics in that author's style. Learned models account for rhythm, rhyme, syntax and semantic constraints.

Other systems do not generate song lyrics independently, but they can assist in the process of writing, as they suggest sentences that suit the metre, the rhythm and rhymes (Abe and Ito, 2012; Shen, 2013). There are also approaches for producing both melody and lyrics (Nakamura and Onisawa, 2009; Toivanen, Toivonen, and Valitutti, 2013). In those, the melody is composed only after producing syntactically-correct lyrics that may also satisfy additional constraints.

3. Background Systems

This section briefly describes the two systems combined in this work: Tra-la-Lyrics, which generates text based on rhythm; and PoeTryMe, a versatile platform for the generation of poetry from a set of seeds and a semantic network.

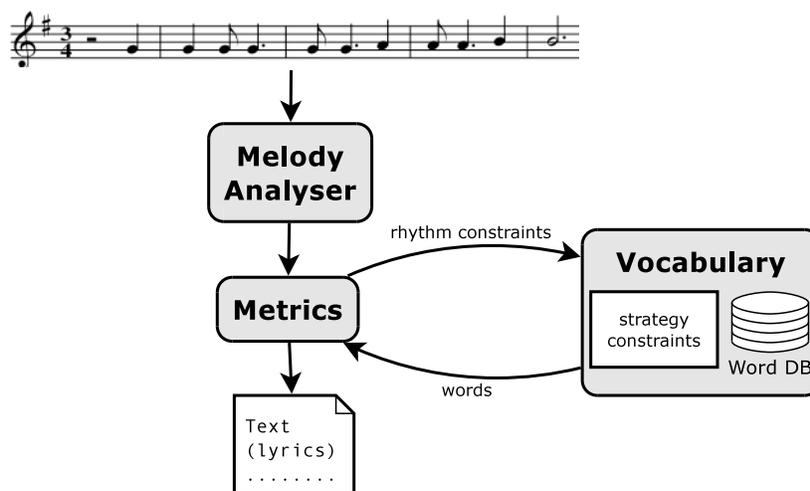


Figure 1: Tra-la-Lyrics architecture (big picture)

3.1 Tra-la-Lyrics

Tra-la-Lyrics generates text automatically, based on a given rhythm, and was originally applied to the generation of song lyrics in Portuguese (Gonalo Oliveira, Cardoso, and Pereira, 2007b,a). Generation is always constrained by the rhythm of a given melody and can be additionally constrained by other features, according to the selected generation strategy. Strategies interact with a relational database, with words and both their rhythm (e.g. syllable division, stress, termination) and morphological features (e.g. POS, number, gender), which is the source of the contents for the generated lyrics.

A shallow view of Tra-la-Lyrics' architecture is displayed in figure 1. The first step towards generation is the analysis of a given melody and its representation as a sequence of beats of different strength levels, which is performed by the Melody Analyser module. Those sequences follow a dot system that sets the metrical accents of each beat inside a metre, and thus the strengths of each note, according to their position (see Lerdahl and Jackendoff (1983)). This process is illustrated in figure 2, for the 4/4 and 3/4 metres. There, each top number represents a crotchet that is divided into four beats (sixteenth). The strength level starts in 0 (the strongest) and grows as the number of dots decreases. Based on the dot representation, the Melody Analyser considers that all beats with a number of dots higher than half the difference between the sum of the highest number and the lowest to be strong, while the others are weak. For example, in the 4/4 metre, strong beats are those with four and three ($4 > \frac{4+1}{2}$, $3 > \frac{4+1}{2}$) dots, namely the first and the third crotchets. In a 3/4 metre, only the beat with three dots, the first, is considered to be strong ($3 > \frac{3+1}{2}$).

After analysing the melody, the separation between the rhythm and the strategy constraints is clear, as there are two core modules: the Metrics module sequentially reads the next beat pattern, sets rhythm constraints accordingly, and passes them to the Vocabulary module, which may add strategy specific constraints. Those, represented in a SQL query to the word database, will define the configuration of the next word to add to the lyrics. Since it is sometimes impossible to satisfy all the constraints, when there are no matching words, some constraints may be dropped or relaxed according to predefined priorities.

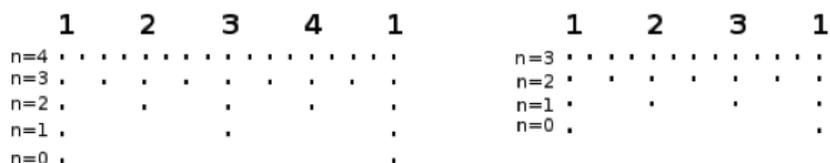


Figure 2: Dot system in 4/4 and 3/4 metres.

The main goal is, nevertheless, that there is a syllable per note. The key for suiting the rhythm is to match stressed syllables with the identified strong beats of the melody. To support the previous intuition, a study was made with 42 Portuguese traditional songs. Based on the dot system, it was confirmed that the majority of the syllables in strong beats were stressed, while in the other beats there were mostly unstressed syllables (see Gonalo Oliveira (2007)). For instance, for the 4/4 metre, 80% of the syllables in the first beat were stressed and 58% of the stressed syllables occurred in the first or third crotchet (strength levels 0 and 1). As for the 3/4 metre, about 87% of the syllables in the first beat were stressed, and these were 57% of all the stressed syllables.

Although several generation strategies were developed, due to the lack of lexical-semantic resources for Portuguese at the time, none of the strategies had in mind the generation of meaningful text, and therefore there were no semantic constraints. In addition to the rhythm constraints, the most simple strategy, Rhythm+Rhymes (R&R), tries to have rhymes in the end of predefined parts of the melody. In another strategy, Generative Grammar (GG), the Vocabulary module adds morphological constraints, so that the selection of words follows the rules of a generative grammar (see Gonalo Oliveira, Cardoso, and Pereira (2007a) for additional details). This may involve backtracking during the generation. Moreover, if a set of seed words is used, the GG strategy will give priority to the selection words with the same dictionary form (lemma, as in *song* and *songs*) as one of the seeds, as long as they match all the other constraints. Given the importance of repetition in song lyrics, both R&R and GG have also a repetition parameter, which may be tuned to increase the number of repeated words.

3.2 PoeTryMe

PoeTryMe (Gonalo Oliveira, 2012; Gonalo Oliveira and Cardoso, 2015) is a generic platform for poetry generation with a flexible architecture. So far, it has been used to produce poetry in Portuguese and Spanish (Gonalo Oliveira et al., 2014), given a set of seed words and different poem structures, including those extracted from song lyrics (Gonalo Oliveira and Cardoso, 2014). Yet, this was done without any deeper notion of rhythm than the organization of lines of specific lengths into stanzas. On the other hand, generated sentences are semantically coherent and use words related to the seeds.

PoeTryMe relies on a modular architecture (see figure 3), which enables the independent development of each module and provides a high level of customisation, depending on the purposes of the user/developer. Different instantiations of PoeTryMe, using alternative implementations of its constituents (e.g. poem structure, transmitted sentiment, semantic network, sentence templates, generation strategy) may thus lead to significantly different poetry generation systems.

The core modules in PoeTryMe’s architecture are the Sentence Generator and the Generation Strategy. The latter parses a poem structure, reads a set of seed words and gets sentences from

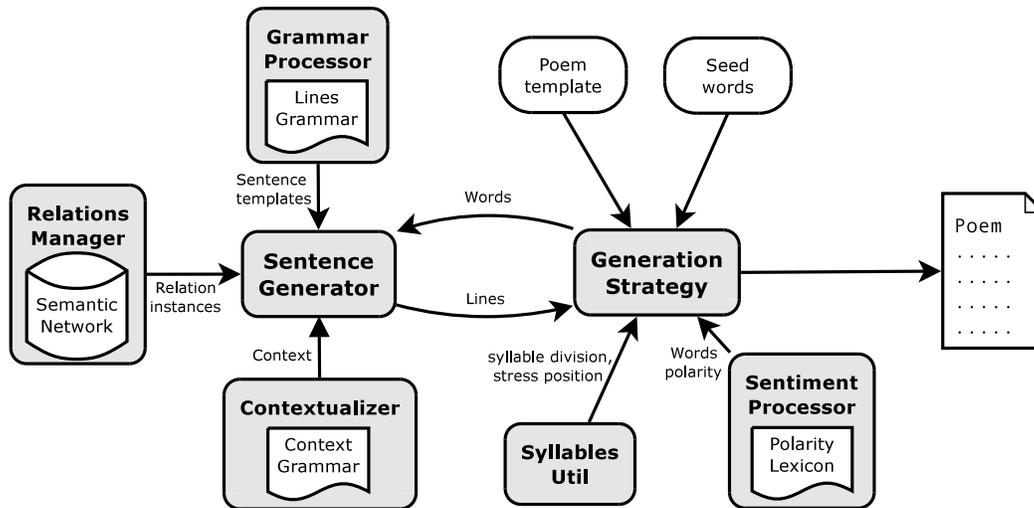


Figure 3: PoeTryMe's architecture

the former. The Sentence Generator produces semantically coherent natural language sentences based on a semantic network and a grammar with line templates. Each template indicates a semantic relation and is filled with a pair of related words, selected from a subgraph of the network, constrained by the seeds. Additional modules, not explored in this work, provide a context to explain the selected lines and words, or constrain generation to words that transmit a predefined sentiment (either positive or negative).

Another feature, not clear in the architecture diagram, is the option of expanding the set of seed words with others that, according to the structure of the semantic network, are relevant. This is performed by applying the Personalized PageRank (Agirre and Soroa, 2009) algorithm to the semantic network, using the given seeds as context (see Gonçalves Oliveira and Cardoso (2014) for additional details).

The Generation Strategy can be reimplemented according to different approaches for exploiting the Sentence Generator and organising retrieved sentences. Although different strategies have been tested, Generate & Test (G&T) is conceptually the most simple and the most explored one. Briefly, for each line in the poem structure, it generates m sentences and selects the best according to a fitness function that considers the target length of the line and, possibly, other features, such as the occurrence of rhymes or alliteration. To increase the probability of rhymes, m is increased according to the order of the line in the stanza. So, for line i , $m_i = m + i.m.\sigma$ sentences are generated, where σ is a predefined parameter.

4. Tra-la-Lyrics 2.0: integrating rhythm in PoeTryMe

Tra-la-Lyrics 2.0 results from the integration of the rhythm component of the original Tra-la-Lyrics in PoeTryMe, thus enabling the implementation of a generation strategy where rhythm is considered. This section describes the integration effort, complemented with illustrative examples of song lyrics generated by the new system.

4.1 Development

The implementation of Tra-la-Lyrics 2.0 involved two parts:

1. Changes in the PoeTryMe architecture, in order to accept a melody as input, instead of a simple poem structure with just stanzas, lines and their lengths;
2. Development of a new generation strategy where the rhythm is considered together with previously considered features.

Both parts relied on the integration of the Melody Analyser module, from Tra-la-Lyrics, in PoeTryMe. Among other features, this enabled us to use the ABC music notation¹ for representing melodies. This notation, used by Tra-la-Lyrics, enables the system to read and write both music and lyrics in ASCII format.

The ABC file provided as input is converted into an internal representation of a stanza with a set of line templates, extended to include the metre type and the sequence of notes, which is later used to identify the sequence of strong and weak beats. Each line template corresponds to a meaningful part of the music. Similarly to the original Tra-la-Lyrics, the division of the music into parts has to be provided by the user, who may select different divisions. In this case, we always opted to split the music into parts that match the sentences of the original lyrics. The end of each part works as a mark for rhymes.

As mentioned earlier, the creation of new generation strategies is straightforward in PoeTryMe. It is a matter of defining the criteria for exploiting the Sentence Generator, including how seed words are manipulated and how generated lines are assessed, selected and organised. To perform generation in Tra-la-Lyrics 2.0, an extension of the original G&T strategy of PoeTryMe was developed with some additional features, to be considered when assessing the suitability of the generated lines. Besides matching the length of the line and the occurrence of rhymes (equation 1), the score function of the implemented strategy (equation 2) also considers the number of unstressed syllables in strong beats, stressed syllables in weak beats and words interrupted by pauses in the melody. Each of the previous features has a customisable weight. Moreover, though they are actually penalties, all the parameter scores are summed to the global score, except for the occurrence of rhymes, which works as a bonus. For historical reasons, in PoeTryMe, the lower the score, the fitter the line is for the target line. Another difference is the way rhymes are considered: instead of restricting it to lines in the same stanza, now there is a three line window between the first and the second line with the same termination. Having three or more lines in the middle of a rhyme would not produce the desired effect.

$$\begin{aligned}
 \text{score}(\text{sentence}, \text{target}, \text{stanza}) = & \\
 & |\#syllables(\text{target}) - \#syllables(\text{sentence})| \\
 & - \alpha * \text{hasRhyme}(\text{sentence}) \\
 & + \beta * \text{lastWordEndsAnother}(\text{sentence}, \text{stanza})
 \end{aligned}
 \tag{1}$$

1. Check <http://abcplus.sourceforge.net/> for more information on the ACB Plus Project.

$$\begin{aligned}
\text{score}(\text{sentence}, \text{melody}) = & \\
& \alpha * |\#\text{syllables}(\text{target}) - \#\text{syllables}(\text{sentence})| \\
& - \beta * \text{hasRhyme}(\text{sentence}) \\
& + \gamma * \text{lastWordEndsAnother}(\text{sentence}, \text{melody}) \\
& + \delta * \#\text{unstressedInStrongBeats}(\text{sentence}, \text{melody}) \\
& + \epsilon * \#\text{stressedInWeakBeats}(\text{sentence}, \text{melody}) \\
& + \zeta * \#\text{interruptionsByPause}(\text{sentence}, \text{melody})
\end{aligned} \tag{2}$$

4.2 Generated Lyrics

The results of Tra-la-Lyrics 2.0 are illustrated with some generated lyrics, together with a rough English translation, for those who do not speak Portuguese. Since music and lyrics are in the ABC notation, it is straightforward to generate a PostScript or PDF score, with a tool like *abcm2ps*². As in the original instantiation of PoeTryMe, in all presented examples, CARTÃO (Gonçalo Oliveira et al., 2011) was used as the semantic network³. We empirically set the parameter $\sigma = 0.75$, and the fitness function parameters: $\alpha = 3$, $\beta = 2$, $\gamma = 1$, $\delta = 1$, $\epsilon = 0.1$, $\zeta = 1$. We admit that they might need some tuning in the future.

In all the presented examples, only one seed word was given, and two other were added based on their relevance, after running the Personalized PageRank in CARTÃO. For instance, for the seed *amor* (love), *sentimentos* (feelings) and *cupido* (cupid) were added, for *investigar* (to investigate), the relevant words were *pesquisar* (to research) and *indagar* (to inquire), and for *música* (music), *instrumentos* (instruments) and *danças* (dances).

Figure 4 displays generated lyrics for the Portuguese traditional song *Alecrim*, which has a 4/4 metre, using the seed *amor*. There is one rhyme (*mão*, *aptidão*) and the rhythm matches the melody closely, as few unstressed syllables are in strong beats. The latter include the stress of *cupidíneos*, and *flotímias*. The sum of the scores of all the lines is 1.9. Regarding the semantics, each line uses one of the words in the extended set of seeds and a semantically-related word in a semantically-coherent sentence. For instance, the word *flotimias* (translated to philotimy), which means love for the honours, is used twice to specify the kind of love.

Figure 5 shows lyrics generated for Johnny Cash’s song *Ring of Fire*, using the seed *investigar*. Similarly to *Alecrim*, this song has a binary division, this time 2/4. The rhythm of the displayed lyrics does not match the melody as closely as in the previous example. There are unstressed syllables in strong beats, such as *pesquisar* or *inquirir*, there is no note for the last syllable of the last word (*ponderar*), and the first word (*sondar*) is interrupted by a pause. This results in a score of 4.2, which is not worse because there are many rhymes. It happens because the seed word, *investigar*, a verb in the infinitive, ends in *-ar*, and there are many other Portuguese verbs with this termination. In this case, the lyrics can be seen as a word play with those verbs, all possible synonyms of *investigar*.

Figure 6 displays lyrics for Bob Dylan’s song *Times They Are a Changin’*, using the seed *música*. This song has a 3/4 metre, which means that, according to the dot system, only the first beat of each metre is strong. Once again, though just a minority, there are words that do not match the rhythm,

2. Check <http://moinejf.free.fr/> for the *abcm2ps* command-line tool.

3. CARTÃO is a lexical-semantic knowledge base, composed by relational triples of the kind (*word_1 predicate word_2*), all extracted from three Portuguese dictionaries.

Alecrim

Tra-la-Lyrics 2.0

sen - ti - men - tos ra - lés ca - la - das a cu - pi - do seu ho - mem fa - vor pe - di - a

cu - pi - dí - ne - os a - mor de mão sen - tí - men - tos não que tu és ap - ti - - dão

eu t'ím - plo - ro a - mor ó meu ser um a - mor de fi - lo - ti - mi - as doi - ra - das

com sen - tí - men - to a - mor su - ão ho - je é a - - mor de fi - lo - ti - mi - a

*sentimentos ralés caladas
a cupido seu homem favor pedia
cupidíneos amor de mão
sentimentos não que tu és aptidão
eu t'imploro amor ó meu ser
um amor de filotimias doiradas
com sentimento amor suão
hoje é amor de filotimia*

feelings of silent anger
to the cupid his man asked a favour
cupidean hand love
no feelings because you are a skill
I beg you love to my being
a love of golden philotimy
with a south wind love feeling
today is philotimy's love

Figure 4: Score of *Alecrim* with lyrics by Tra-la-Lyrics 2.0, using the seed *amor*.

including *reparte*, *para*, *pipocas* or *perdendo*. There are also two rhymes (*menor/justador* and *refrão/transmissão*), which leads to an overall score of 3.5. On the semantics, despite the presence of words strongly-associated to music (e.g. *refrão* or *danças*), there are several less clear and indirect connections. For instance, *libertista* and *justador* are kinds of person, which is a hypernym of musician. The last line has a more distant connection, produced due to the ambiguity of the Portuguese word *papa*, which might mean *pope* or *thick food*. The pope meaning is related to gregorian music, but the used word, *xarém*, is actually a kind of Portuguese food. While this kind of ambiguities can lead to odd connections, sometimes they add a new dimension to the text, close to the (unintentional) use of figurative language.

5. Validation of the Results

The section reports on three experiments towards the validation of Tra-la-Lyrics 2.0. To analyse the main differences and confirm if there were improvements, we compared lyrics by Tra-la-Lyrics 2.0 with those by with the original Tra-la-Lyrics, under similar conditions. More precisely, lyrics were generated with the most simple, R&R, and the GG strategies of Tra-la-Lyrics, both with 35% on the word repetition parameter, and with source words from a large coverage morphological lexicon of Portuguese.

Ring of Fire

Tra-la-Lyrics 2.0

son - - dar e in - da - gar in - da - gar e al - vei - tar

ler e des - co - - brir fu - ro - ar e in - ves - ti - - gar

sem - pre a bas - cu - lhar e a pes - qui - sar in - da - gar em fren - te ir a in -

qui - rir des - co - - brir e ler o - lhar e pon - derar

sondar e indagar probe and inquire
indagar e alveitar inquire and ascertain
ler e descobrir read and discover
furoar e investigar ferret and investigate
sempre a basculhar e a pesquisar always rummage and search
indagar em frente ir a inquirir ask ahead go inquire
descobrir e ler find and read
olhar e ponderar look and consider

Figure 5: Score of *Ring of Fire* with lyrics by Tra-la-Lyrics 2.0, using the seed *investigar*.

All the experiments used the same datasets for comparison, here described before the experiments, right after this paragraph. Each experiment is focused on one of three features, which we believe to be key in song lyrics, namely: (i) rhythm; (ii) rhymes; and (iii) semantics. We try to show that, although there are less rhythm matches than in the simplest strategy, this loss is not extreme, especially if we consider that the text now follows a grammatical structure and uses words associated to the given seeds, in semantically-coherent sentences.

5.1 Generated Datasets

We used each selected generation strategy to produce a set of lyrics for the same songs and using the same seeds. The same dataset was used for all the performed experiments. Since it is impossible to compare the results in every possible scenario, we tried to cover different melodies, of different metres, as well as different seed words. More precisely, for each combination of a strategy, a melody and a Portuguese seed word (when applicable), five lyrics were generated. As mentioned earlier, the strategies were the following three:

- Tra-la-Lyrics Rhythm & Rhymes (R&R)
- Tra-la-Lyrics Generative Grammar (GG)

Times they are a changin'

Tra-la-Lyrics 2.0
Bob Dylan

pes - so - a li - ber - tis - ta em dó me - nor re - par - te a pes - so - a e o jus - ta
dor ins - tru - men - tos de ten - ta pa - ra o re - frão é pre - ci - so dan - ças pi - po - cas
per - den - do se as es - fe - ras da trans - mis - são o que foi pa - pa de xa - réns

*peessoa libertista em dó menor
reparte a pessoa e o justador
instrumentos de tenta para o refrão
é preciso danças pipocas
perdendo-se as esferas da transmissão
o que foi papa de xaréns*

libertarian person in C minor
redistributes the person and the fighter
instrument of probe to the chorus
it takes dances popcorn
if lost the transmission domains
what was xarém food

Figure 6: Score of *Times They Are a Changin'* with lyrics by Tra-la-Lyrics 2.0, using the seed *música*.

- Tra-la-Lyrics 2.0

The selected seeds included one of the most common subjects in poetry (love) plus five seed words related to the topic of this work, more precisely, the following five nouns and one verb:

- *amor* (love)
- *poesia* (poetry)
- *música* (music)
- *cantiga* (song)
- *ciência* (science)
- *investigar* (to research)

Since R&R does not accept a seed word, the dataset for that strategy is six times smaller. The melodies used in these experiments cover two of the most common metre types, namely those with a binary division (2/4 and 4/4) and those with a ternary division (3/4). Two of those melodies were used during the development of the original Tra-la-Lyrics, namely:

- *Alecrim* (4/4), Portuguese traditional song
- *Um Barquinho* (3/4), Portuguese traditional song

Four additional melodies of well-known pop-rock songs, with original lyrics in English, were also used:

- *Bad Moon Rising* (4/4), by Creedence Clearwater Revival
- *Ring of Fire* (2/4), by Johnny Cash
- *House of the Rising Sun* (3/4), performed by The Animals (among others)
- *Times they are a Changin'* (3/4), by Bob Dylan

The complete dataset included 390 song lyrics, 30 generated with R&R, 180 by the GG strategy, and 180 by Tra-la-Lyrics 2.0.

5.2 Rhythm validation

As confirmed in a study made in the scope of Tra-la-Lyrics (Gonçalo Oliveira, 2007), song lyrics have roughly a syllable matching each note of the melody. Moreover, there is a strong correlation between the stresses in the text and the strong beats of the melody. More precisely, the majority of notes in strong beats match stressed syllables, while unstressed syllables match weaker beats. Therefore, to validate the rhythm of the generated lyrics, we used a tool to analyse the distribution of stressed and unstressed syllables, according to the metre type – the higher the number of stressed syllables matching strong beats and the number of unstressed syllables in weak beats, the closer the rhythm is matched. The tool used for this purpose was originally developed for the validation of Tra-la-Lyrics and relies on the dot system (see section 3.1) to identify the strength of each note. For syllable division, termination extraction and stress identification, it uses the same module as Tra-la-Lyrics and as the Portuguese version of PoeTryMe.

Figures 7 and 8 show the distribution of stressed and unstressed syllables in the beats of the binary and ternary metres, respectively in the lyrics by the R&R and the GG strategies. There is a pair of bars for each note strength level, which can go from 0 (strongest) to 3 (weakest). In the 4/4 metre, level-0 and level-1 beats are considered strong, while in the 3/4, only the level-0 are. The charts show that, with both R&R and GG, more than 80% of the syllables in a 0-level beat of a binary metre are stressed. This number is 78% or higher for 3/4. On the weakest levels, stressed syllables are scarce, which satisfies the main goal of Tra-la-Lyrics.

Figure 9 shows the same numbers, but for Tra-la-Lyrics 2.0. Although there are 9-14% more unstressed syllables in strong beats, and more stressed syllables in weak beats, the trend is still clearly to have more stresses in strong beats ($\approx 70\%$ for 4/4 and $\approx 79\%$ for 3/4). Therefore, we can say that, although not as closely as for the other strategies, the rhythm is still matched.

5.3 Rhymes

Though not mandatory, rhymes are a common feature in poetry and music. They are relevant for setting the rhythm of the text and are often attached to the poetic form. However, more than a mere matching of stresses, rhymes also involve the repetition of the same sound, typically in the end of two lines close to each other. Therefore, they are assessed as a specific feature. We consider that there is a rhyme every time two lines, in a window of four or less, have the same termination. Although there are different kinds of rhyme, only the most consensual type for Portuguese was

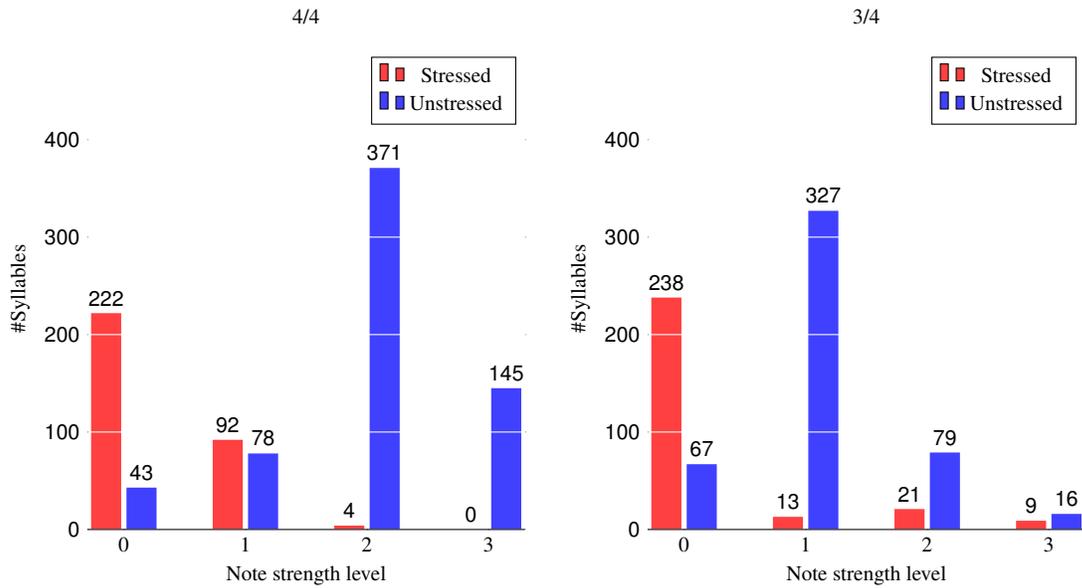


Figure 7: Matches between syllable types and beat strength level according to metre type with the R&R strategy.

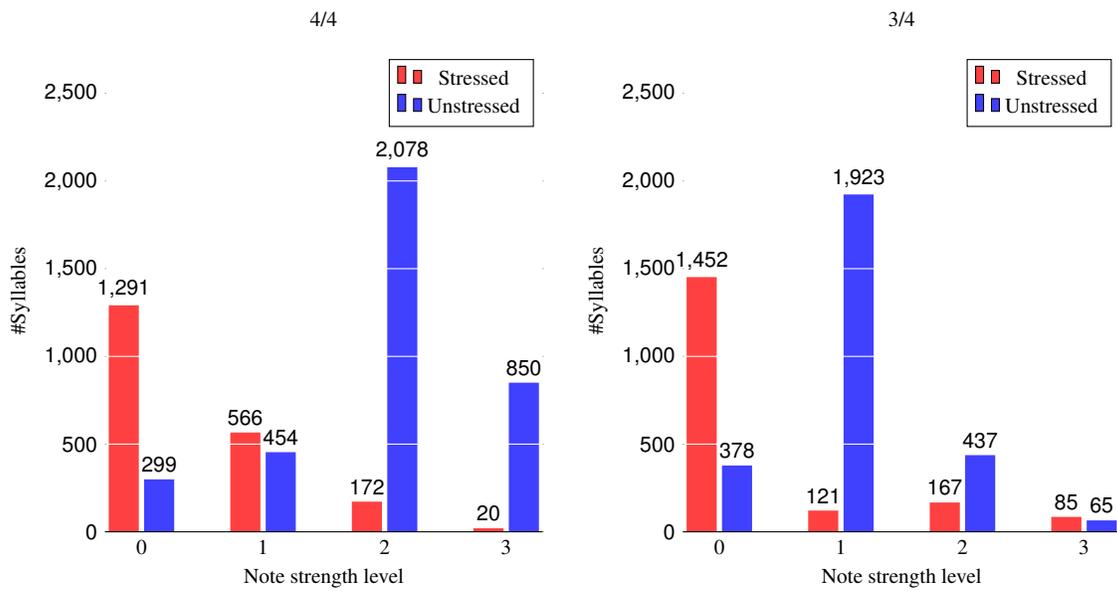


Figure 8: Matches between syllable types and beat strength level according to metre type with the Generative Grammar strategy.

targeted: the last words of the rhyming lines must have the same stressed vowel and the same sound from that place to the end of the word, often obtained by the same sequence of letters.

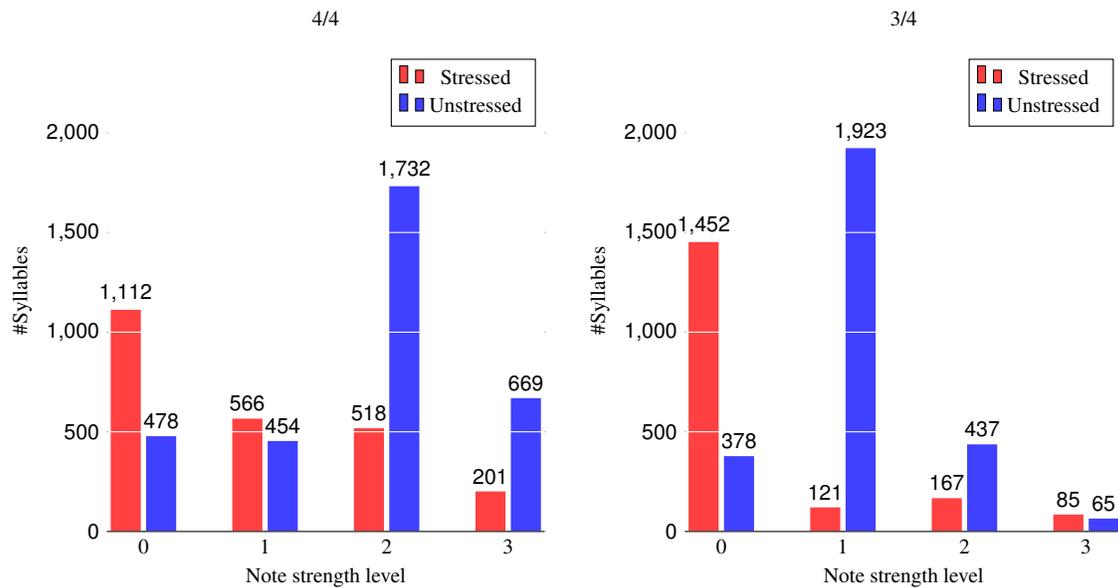


Figure 9: Matches between syllable types and beat strength level according to metre type with PoeTryMe.

Given the role that rhymes play, we compared the ratio of rhyming lines in the lyrics generated by each of the three strategies. The results are displayed in table 1, first according to the melody, and then to the used seed word. Although rhymes are more frequent with the R&R strategy (60-72% of the lines), Tra-la-Lyrics 2.0 generated, on average, more rhymes than the GG strategy. But the presence of rhymes is more consistent in the latter, while in Tra-la-Lyrics 2.0 they are more dependent on the seed. For instance, with the word *investigar* (and possibly any other verb ending with *-ar* that has synonyms), rhymes are much more common and occur in about 88% of the lines. An example of this situation was displayed in figure 5. Using the same strategy and the seed *música*, only 17% of the lines rhyme. This is the only seed that leads to less rhymes than the GG strategy. No other conclusions can be taken about melodies or seed words that lead to more or less rhymes.

Song	#Lines	R&R	GG	2.0
<i>Alecrim</i>	8	0.60	0.21	0.51
<i>Bad Moon</i>	7	0.63	0.21	0.30
<i>Ring of Fire</i>	8	0.65	0.34	0.43
<i>Barquinho</i>	5	0.72	0.38	0.31
<i>House</i>	4	0.70	0.11	0.28
<i>Times</i>	6	0.60	0.31	0.52

Seed	GG	2.0
<i>amor</i>	0.26	0.37
<i>música</i>	0.31	0.17
<i>cantiga</i>	0.25	0.31
<i>poesia</i>	0.24	0.34
<i>ciência</i>	0.28	0.37
<i>investigar</i>	0.25	0.88

Table 1: Ratio between number of lines and rhymes in the lyrics generated with the different strategies, for each melody and for each seed word.

5.4 Semantics

Of the three selected features, semantics is the most difficult to evaluate, and also the most subjective. Poetry typically exploits linguistic features such as metaphors or figurative language, and not always transmits a well-defined message, either on purpose or not. Moreover, the occurrence of other features, including the regular metre, rhymes or repetition, sometimes compensate for the apparent lack of sense.

In this section, before presenting some hints on the semantics of the generated lyrics, we discuss the (lack of) semantics in the original Tra-la-Lyrics and present PMI, the measure we applied for quantifying the semantic similarity between the words of the generated lyrics used and the used seed.

5.4.1 SEMANTICS IN THE ORIGINAL TRA-LA-LYRICS

To give an idea on the kind of lyrics by the original Tra-la-Lyrics, figures 10 and 11 present two lyrics for the song *Times They Are a Changin'*, generated respectively by the R&R and the GG strategies. The same figures show a rough translation of those lyrics, which was quite difficult and sometimes relied on the most common translation because, here, the context does not really matter. We expect to illustrate that, in R&R, despite some repetitions and rhymes, the sequences are random, while in GG lines follow (more or less) grammatical sequences. Also, more than simple repetitions, GG tries to use words with the same lemma as the given seed. In the example, the seed was *música* and there are occurrences of this word in the diminutive singular (*musiquita*) and plural (*musiquitas*). If the seed were a verb, there would be different conjugations.

<i>esmoutes aptou contrafecho zavaís</i>	clear adapted voussoir bites
<i>coserás esguiáramo curta grifais</i>	will sew have wounded short italics
<i>esguiáramo trias queijasses bedum</i>	have slendered selected would make cheese bad smell
<i>esguiáramo abetum um</i>	have slendered filling one
<i>escambrões esguiáramo sodra aptou</i>	unsociable have slendered stifle joint adapted
<i>esguiáramo zavaís um</i>	have slendered bites one

Figure 10: Lyrics generated by the Rhythm & Rhymes strategy for the song *Times They Are a Changin'*

<i>noventa candis permaneçam um flash</i>	ninety lamps remain a flash
<i>raziava pelara cintilem milhões</i>	devastated had frayed millions shimmer
<i>casibeques hariolo roques camões</i>	shot coat diviner rooks camões
<i>desasavam nonos belhós nos</i>	unroasted ninth fritter in the
<i>ratinado trezentas musiquitas mil</i>	frizzed three hundred little songs a thousand
<i>musiquita olivar pui</i>	little song olivary fray

Figure 11: Lyrics generated by the Generative Grammar strategy for the song *Times They Are a Changin'*, using the seed *música*

Among other questions, in a survey performed to evaluate Tra-la-Lyrics (Gonçalo Oliveira, 2007; Gonçalo Oliveira, Cardoso, and Pereira, 2007a), 70 human volunteers were asked to give a suitable title (if possible) and to rate the meaning transmitted by lyrics generated by different strategies in a scale between 0 (nonsense) to 4 (clear meaning). Although none of the strategies had

any concern with the semantics, the majority of the raters could give their own title for the lyrics, often including one of the used words. As expected, the meaning was always rated negatively but, for the GG strategy the average rating was 1.66, much higher than the 0.57 for the R&R strategy. This suggests that, though not enough, it is easier to transmit a meaning if a grammatical structure is followed.

5.4.2 POINTWISE MUTUAL INFORMATION

Though we could see the aforementioned numbers as a baseline for the quality of semantics in a system where it is not handled, we first approached the assessment of this feature with Pointwise Mutual Information (PMI), a statistical measure typically used to compute distributional similarity/word association. The PMI of a pair of outcomes (e.g. words a and b) quantifies the discrepancy between the probability of their coincidence (e.g. co-occurrence, $P(a, b)$) given their joint distribution and their individual distributions ($P(a)$ and $P(b)$), assuming their independence. Mathematically, PMI can be computed according to equation 3, where we added that probabilities can be obtained from the number of hits of the analysed words in a search engine.

$$PMI(a, b) = \log \frac{P(a, b)}{P(a) * P(b)} = \log \frac{Hits(a \cap b)}{Hits(a) * Hits(b)} \quad (3)$$

On the semantic similarity domain, PMI has successfully been used to identify synonyms (Turney, 2001), or to evaluate the coherence of topic models (Newman et al., 2010), among others. In both the previous cases, it has been noticed that PMI scores had a high correlation with those by human judges. When measured on a large corpus, pairs of words typically used together, and rarely alone (semantically similar), have higher PMI values than those that do no co-occur or only co-occur by accident.

In this work, PMI was measured in the Portuguese Wikipedia, which is a large corpus, with a large coverage, and created independently from our line templates and semantic network. Moreover, it provides a REST API⁴ that eases the process of counting the number of documents where a word or a pair of words occur. To use only the content words, stopwords were removed based on the Snowball list for Portuguese⁵.

The higher the PMI, the higher association between the word pairs. To give a closer insight on this measure, in table 2, we reveal some high PMI values between used seeds and another word. In opposition to those, pairs with low PMI scores include, for instance, {poesia (poetry), fracções (fractions)} (-1.372), {música (music), plantas (plants)} (-1.064), or {ciência (science), canção (song)} (-0.657). When the words of the pair never co-occurred, or when one of the words was not in Wikipedia, PMI was set to 0.

5.4.3 RESULTS

To have some hints on the semantic similarity between all the words used in the lyrics and also between these and their seeds, we measure the average PMI between every distinct content word used and:

- The seed word (only applies for GG and 2.0) – $\overline{PMI}(\text{seeds})$.

4. Check <http://pt.wikipedia.org/w/api.php>

5. Check <http://snowball.tartarus.org/algorithms/portuguese/stop.txt>

a	b	$PMI(a, b)$
<i>cantiga</i> (song)	<i>trova</i> (ballad)	4.868
<i>cantiga</i> (song)	<i>embalar</i> (to rock)	4.556
<i>investigar</i> (investigate)	<i>interrogar</i> (ask)	4.306
<i>poesia</i> (poetry)	<i>poético</i> (poetic)	3.993
<i>investigar</i> (investigate)	<i>averiguar</i> (ascertain)	3.621
<i>poesia</i> (poetry)	<i>musa</i> (muse)	2.566
<i>ciência</i> (science)	<i>tecnologias</i> (technologies)	2.176
<i>música</i> (music)	<i>refrão</i> (chorus)	2.113
<i>música</i> (music)	<i>melodia</i> (melody)	2.041
<i>ciência</i> (science)	<i>educação</i> (education)	1.908
<i>amor</i> (love)	<i>mulher</i> (woman)	1.726
<i>amor</i> (love)	<i>paz</i> (peace)	1.631

Table 2: Examples of highly associated words according to their PMI in the Portuguese Wikipedia.

- Every other content word used – $\overline{PMI}(all)$.

In addition to those measures, the ratio between distinct content words used and all content words in each generated lyrics was computed. This should give an idea on the variation of the lyrics by each strategy. If this ratio is close to 1, there are few repetitions, which means that novelty is high but consistency might be low. On the other hand, if it is close to 0, there are consistently repeated words, which is not very interesting. Not to mention that this number can bias the standard deviation of the average PMI. We can say we are looking for a middle-ground.

For all lyrics generated with the R&R strategy, the distinct words ratio was 0.995, which shows that few words are repeated. Although we set a repetition parameter to 35%, this constraint is frequently dropped because it has a low priority. For the same lyrics, as expected, $\overline{PMI}(all)$ had a low value, more precisely, an average of 0.585 ± 1.892 . Tables 3 and 4 show the result of the previous measures according to the seed word used, respectively for the strategy GG and Tra-la-Lyrics 2.0. The latter has more repetitions, which are either the seeds or related words, and GG suffers from a similar problem as R&R, in this measure. The tables also show that the PMI values are higher for GG than for R&R, but this only happens because the former tries to use words with the same lemma as the seed, which have a high association. The value of both PMI measures for every used seed is clearly higher for Tra-la-Lyrics 2.0, which confirms that we are heading towards the right direction, when it comes to generating song lyrics on a semantic domain.

But PMI only accounts for distributional semantics, a weak sense of semantics. Although a high PMI value is not enough to confirm that the generated lyrics are meaningful and that the whole poem makes sense, it shows clear improvements on this aspect, when compared to the original Tra-la-Lyrics. Moreover, these values should be seen in the context of the whole system architecture. It should be noted that PoeTryMe not only relies on a set of syntactically-correct line templates, but also that each template transmits a known semantic relation, where blanks are filled with words that, according to the semantic network, are related by the template relation. This enables the production of semantically-coherent sentences which, together with the restrictions on the network, set by the seeds, also guarantees the selection of words of the same semantic domain. On the other hand, some templates might contain additional content words but, according to the presented results, they do not seem to have a noticeable impact on the overall semantics.

Seed	#Words	(distinct)	$PMI(seeds)$	$PMI(all)$
<i>amor</i>	603	(0.923)	1.149 ±1.092	0.966 ±1.845
<i>música</i>	582	(0.928)	1.109 ±1.760	1.017 ±1.968
<i>cantiga</i>	606	(0.968)	0.661 ±0.582	0.952 ±1.947
<i>poesia</i>	577	(0.922)	1.417 ±1.600	1.099 ±1.982
<i>ciência</i>	575	(0.918)	0.697 ±0.557	0.949 ±1.886
<i>investigar</i>	571	(0.923)	1.082 ±1.229	1.116 ±1.987

Table 3: Distinct words and PMI for the lyrics generated by the GG strategy, according to the given seed.

Seeds	#Words	(distinct)	$PMI(seeds)$	$PMI(all)$
<i>amor</i>	473	(0.727)	1.506 ±1.273	1.397 ±1.710
<i>música</i>	433	(0.623)	2.253 ±2.467	1.675 ±2.070
<i>cantiga</i>	397	(0.740)	1.164 ±0.843	1.547 ±1.761
<i>poesia</i>	451	(0.832)	1.554 ±1.315	1.437 ±1.620
<i>ciência</i>	483	(0.785)	0.846 ±0.772	1.297 ±1.764
<i>investigar</i>	369	(0.716)	1.314 ±1.120	1.221 ±1.390

Table 4: Distinct words and PMI for the lyrics generated with Tra-la-Lyrics 2.0, according to the given seed.

5.5 Crowdsourced evaluation

All the previous evaluations are quite objective and rely on predefined metrics. In order to get more personal opinions on the lyrics, including their most subjective features, the evaluation datasets were uploaded to the Crowdfunder platform⁶, together with a survey set to be answered by Portuguese contributors, living in Portugal. For each piece of lyrics, the name of the original song was provided with a link for one of its versions in Youtube, with the original lyrics. Instructions stated that, before start answering, each contributor would have to listen to the original song and then try to sing the displayed lyrics with its rhythm. Each assignment contained three lyrics and we made sure that the same contributor could not answer more than three assignments. Also, in order to have the same proportion of lyrics produced by all the strategies, each of those by R&R was uploaded six times.

Though slightly different, the survey was inspired by the one made in the evaluation of the original Tra-la-Lyrics. It had the following seven questions:

1. **Rhythm:** how well does the rhythm of the text suit the original melody? Is it possible to sing it as the lyrics of that melody?
 - Poorly (1) to Perfectly (5)
2. **Rhymes:** when some lines in the lyrics have a similar termination, a rhyme occurs. What is the quality of the rhymes in the displayed lyrics?
 - No rhymes (1) to Excellent (5)

6. <http://www.crowdfunder.com/>

3. **Sound:** regardless of the original melody, do the words of the displayed text have a pleasant sound?
 - Very unpleasant (1) to Very pleasant (5)
4. **Grammar:** how well do the sentences follow the main grammatical rules?
 - Poorly (1) to Perfectly (5)
5. **Meaning:** does the text make sense? Is it possible to, somehow, interpret it?
 - Nonsense (1) to Perfectly clear sense (5)
6. **Subject:** is the text related to any of the following subjects?
 - *amor, música, cantiga, poesia, ciência, investigar* or none of the previous
7. **General quality:** in general, what is the quality of the displayed text when combined with the music?
 - Poor (1) to Excellent (5)

In the end of this evaluation, a total of 287 surveys were answered, more precisely for 94 lyrics by R&R, 96 by GG and 97 by Tra-la-Lyrics 2.0. The average scores, standard deviations, and median by strategy are displayed in table 5.

Feature	R&R		GG		2.0	
	Means \pm SD	Median	Means \pm SD	Median	Means \pm SD	Median
Rhythm	2.564 \pm 1.151	3	2.589 \pm 1.087	3	2.742 \pm 0.992	3
Rhymes	2.755 \pm 1.094	3	2.421 \pm 1.116	3	2.784 \pm 1.073	3
Sound	2.830 \pm 1.132	3	2.684 \pm 1.160	3	3.010 \pm 1.036	3
Grammar	2.511 \pm 1.105	2	2.379 \pm 1.141	2	2.856 \pm 1.090	3
Meaning	2.287 \pm 1.151	2	2.379 \pm 1.178	2	2.732 \pm 1.075	3
Subject (= <i>seed</i>)	0.564 \pm 0.499	-	0.137 \pm 0.346	-	0.206 \pm 0.407	-
Subject none	0.564 \pm 0.499	-	0.411 \pm 0.495	-	0.351 \pm 0.480	-
General	2.521 \pm 1.133	3	2.463 \pm 1.137	3	2.680 \pm 1.095	3

Table 5: Results of the crowdsourcing evaluation.

No strong conclusions can be taken from the results, because they are very tight. All average scores fall in the [2,3] interval, except the sound in Tra-la-Lyrics 2.0, which got an average score of 3.01. Nevertheless, Tra-la-Lyrics outperformed the other strategies, not only on the meaning evaluation but, surprisingly, also on all the other questions but one: for R&R, more than half of the contributors selected *none of the previous* as the subject, which was correct, because this strategy is not based on a seed word. It is lower than one would expect, but still clearly higher than the 20% of contributors that selected the correct seed for Tra-la-Lyrics 2.0. We should still recall that there were some semantically-close seeds, such as *music* and *song*, or *science* and *research*. At first, it might seem strange that the rhythm evaluation is higher for Tra-la-Lyrics 2.0 than for R&R. Our intuition is that, even though some words might not match the rhythm, the independent generation of full lines for each part of the music helps to keep a coherent pace. Moreover, each line is produced from a template, collected from human-created poetry, which might have favoured the average

scores of the sound feature in Tra-la-Lyrics 2.0. The difference on the rhymes of the previous strategies is not significant. Although still close, the meaning feature is where the differences are more pronounced, which makes sense, since one of our goals was to go one step further towards the production of semantically-coherent lyrics. But it is still 2.7, on average, just slightly higher than the middle of the scale.

Another less expected result is the grammar feature, which is better for R&R than for GG. This leads us to some of the limitations of this evaluation that should be acknowledged. For instance, it relied on the opinion of non-expert contributors, who probably do not have a strong knowledge on grammar or music. Also, we noticed that not every contributor was aware of what a rhyme is, as we found a situation where, although there were no rhymes, this feature was scored 5. When detected, this kind of situation lead to the exclusion of all the answers by the same contributor.

Finally, even with the link to the original song, there was no guarantee whether the contributors would fit the lyrics in the original rhythm properly, if they would spend much time trying to figure it out or if they would be satisfied the first time they sang the lyrics. When a similar evaluation was made for Tra-la-Lyrics, an interpretation of the generated lyrics, with voice and guitar, was provided with each survey. Back then, it was possible, because the survey targeted just one music and three generated lyrics. This time, it was impractical to play and record 390 lyrics. In the future, we might explore some voice synthesis tool for singing the produced lyrics.

Figure 12 displays one of the lyrics that got the best scores in this evaluation: Rhythm=4, Rhymes=3, Sound=5, Grammar=4, Meaning=4, Subject=*amor* (correct), General=4. The lyrics were produced by Tra-la-Lyrics 2.0 for the song *Alecrim*.

<i>ranchos são a uma pessoa</i>	healthy swarms to a person
<i>dei o teu deus à marcha do meu cupido</i>	I gave your god to the march of my cupid
<i>mulher cupido em dó menor</i>	cupid woman in C minor
<i>as pessoas não têm elaborador</i>	people do not have a manufacturer
<i>eu t'imploro amor ó meu ser</i>	I beg you love my living thing
<i>e com seus pirralhos e mulher se parte</i>	and with their brats and woman breaks herself
<i>sem atrábilis nem cólera</i>	without irritation nor anger
<i>nos fortes pirralhos na grande mulher</i>	in the strong brats in the great woman

Figure 12: Lyrics generated by Tra-la-Lyrics 2.0 for the song *Alecrim*, using the seed *amor*

6. Concluding remarks

Matching the rhythm very closely and using features such as rhymes and other kinds of repetition might be enough to produce song lyrics with a great sound. And the sounds do not even have to be words, as in scat music. However, when lyrics transmit a meaningful message, a completely new dimension is given to the song. This is why most popular songs have linguistically well-constructed lyrics, in terms of syntax and semantics, even if their interpretation is not always obvious.

Tra-la-Lyrics generates text with word stresses matching the rhythm of a given melody. Produced lyrics might contain additional features, such as occasional rhymes, repeated words, or even follow syntactically-correct sequences. But there are not any concerns on semantics, which leads to lyrics with an odd selection words and a semantically-incoherent message.

A new version of Tra-la-Lyrics, with a semantic component, was presented here. It combines the rhythm handling modules of Tra-la-Lyrics with PoeTryMe, a platform for the generation of poetry on a semantic domain, defined by a set of seed words and a semantic network. This was made

possible due to the flexible architecture of PoeTryMe and the adaptation of a generation strategy to encompass the assessment of the rhythm.

After a description of the effort involved in the creation of Tra-la-Lyrics 2.0, several features were analysed to assess the improvements towards the previous version of this system. It focused on three main features, namely rhythm, rhymes and the semantic similarity of the words in the lyrics generated by this system. Despite less syllables following the rhythm, the trend is still to have more stressed syllables matching the strong beats. Rhymes occur less than in the simplest strategy, but more than when a generative grammar is used. But when it comes to semantics, Tra-la-Lyrics 2.0 uses words with a higher semantic association among them and with the generation seeds.

Besides the automatic analysis, a survey, answered by several contributors, confirmed that the lyrics of Tra-la-Lyrics 2.0 show improvements towards those by the original system. Despite close results on different features, Tra-la-Lyrics 2.0 got the highest scores, not only on the meaning and sound but, surprisingly, also on the rhythm and rhymes.

Even if the previous are not enough to claim that Tra-la-Lyrics 2.0 generates meaningful lyrics, together with our knowledge of the system, the obtained results are positive, and show that we are one step further towards that goal. But there is still much room for improvement. For instance, some parameters should be tuned, starting with the number of generations per line, currently 1000. Increasing this number would lead to more fit lines in a trade-off for more generation time, currently about 30 seconds for a poem with six lines. Moreover, other generation strategies could be applied, including adaptations of those already implemented for PoeTryMe (e.g. an evolutionary algorithm). Another improvement could result from collecting more line templates. Currently, although some templates were learned from original song lyrics in Portuguese, the majority came from written poetry. Further results can be compared with the current, the same way the current were compared with those by the original Tra-la-Lyrics. This is especially true for the results of the automatic analysis, as they do not rely on human opinions. In the future, additional tests should be made to check, for instance, if the results hold for melodies with composed metres (e.g. 3/8, 6/8), and not just those with binary and ternary divisions.

Before concluding, it should be mentioned that, although this work has been applied to Portuguese, as long as there are language specific resources (semantic network, poetry/lyrics collection, morphological lexicon, and syllable division tool), PoeTryMe can be used to generate poetry in other languages. In fact, after a collaboration with researchers of the Universidade Complutense de Madrid, Spain, PoeTryMe currently generates poetry in Spanish (Gonçalo Oliveira et al., 2014) and English, which means that Tra-la-Lyrics 2.0 can also generate song lyrics in those languages. To illustrate the English version of Tra-la-Lyrics 2.0, we end this article with figure 13, which displays selected English lyrics, produced by this system for Bob Dylan's *Times they are a Changin'*, using "software" as the seed word.

Acknowledgments

This work was developed in the scope ConCreTe – *Concept Creation Technology*. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733.

*instead of performances program shall be
and loop nor coterie nor the billowing sea
by computing technologies that onward creep
these functions of my package I keep
with computing technologies every eye
in computing there is AI*

Figure 13: English lyrics, produced by Tra-la-Lyrics 2.0 for Bob Dylan’s *Times they are a Changin’*, using the seed word “software”.

References

- Abe, C., and Ito, A. 2012. A Japanese Lyrics Writing Support System for Amateur Songwriters. In *Proceedings of 4th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2012*.
- Agirre, E., and Soroa, A. 2009. Personalizing PageRank for word sense disambiguation. In *Proceedings of 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL’09*, 33–41. Athens, Greece: ACL Press.
- Agirrezabal, M.; Arrieta, B.; Astigarraga, A.; and Hulden, M. 2013. POS-Tag Based Poetry Generation with WordNet. In *Proceedings of the 14th European Workshop on Natural Language Generation*, 162–166. Sofia, Bulgaria: ACL Press.
- Barbieri, G.; Pachet, F.; Roy, P.; and Esposti, M. D. 2012. Markov Constraints for Generating Lyrics with Style. In *Proceedings of 20th European Conference on Artificial Intelligence, ECAI 2012*, 115–120. Montpellier, France: IOS Press.
- Binsted, K., and Ritchie, G. 1994. An Implemented Model of Punning Riddles. In *Procs. of 12th National Conf. on Artificial Intelligence (Vol. 1), AAAI ’94*, 633–638. Menlo Park, CA, USA: AAAI Press.
- Chrismartin, B.; Tobing, L.; and Manurung, R. 2015. A chart generation system for topical metrical poetry. In *Proceedings of the 6th International Conference on Computational Creativity, Park City, Utah, USA, ICC 2015*.
- Colton, S., and Wiggins, G. A. 2012. Computational Creativity: The Final Frontier? In *Proceedings of 20th European Conference on Artificial Intelligence (ECAI 2012)*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, 21–26. Montpellier, France: IOS Press.
- Colton, S.; Goodwin, J.; and Veale, T. 2012. Full FACE poetry generation. In *Proceedings of 3rd International Conference on Computational Creativity, Dublin, Ireland, ICC 2012*, 95–102.
- Gervás, P. 2001. An expert system for the composition of formal Spanish poetry. *Journal of Knowledge-Based Systems* 14:200–1.
- Gervás, P.; Díaz-Agudo, B.; Peinado, F.; and Hervás, R. 2005. Story plot generation based on CBR. *Knowledge-Based Systems* 18(4):235–242.

- Gonçalo Oliveira, H., and Cardoso, A. 2014. Using a generic poetry generation system to produce song lyrics with sentiment. In *Proceedings of 3rd Workshop on Computational Creativity, Concept Invention, and General Intelligence (C3GI@ECAI 2014)*, volume 01-2014 of *Publications of the Institute of Cognitive Science, Osnabrück*. Prague, Czech Republic: Institute of Cognitive Science.
- Gonçalo Oliveira, H.; Antón Pérez, L.; Costa, H.; and Gomes, P. 2011. Uma rede léxico-semântica de grandes dimensões para o português, extraída a partir de dicionários electrónicos. *Linguamática* 3(2):23–38.
- Gonçalo Oliveira, H.; Hervás, R.; Díaz, A.; and Gervás, P. 2014. Adapting a Generic Platform for Poetry Generation to Produce Spanish Poems. In *Proceedings of 5th International Conference on Computational Creativity, Ljubljana, Slovenia, ICC 2014*.
- Gonçalo Oliveira, H. R.; Cardoso, F. A.; and Pereira, F. C. 2007a. Exploring different strategies for the automatic generation of song lyrics with Tra-la-Lyrics. In *Proceedings of the 13th Portuguese Conference on Artificial Intelligence, EPIA 2007*, 57–68. Guimarães, Portugal: APPIA.
- Gonçalo Oliveira, H. R.; Cardoso, F. A.; and Pereira, F. C. 2007b. Tra-la-Lyrics: an approach to generate text based on rhythm. In *Proceedings of 4th International Joint Workshop on Computational Creativity*, 47–55. London, UK: IJWCC 2007.
- Gonçalo Oliveira, H. R. 2007. Geração de texto com base em ritmo. Master's thesis, University of Coimbra.
- Gonçalo Oliveira, H. 2012. PoeTryMe: a versatile platform for poetry generation. In *Proceedings of ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence, Montpellier, France, C3GI 2012*.
- Gonçalo Oliveira, H., and Cardoso, A. 2015. Poetry generation with PoeTryMe. In Besold, T. R.; Schorlemmer, M.; and Smail, A., eds., *Computational Creativity Research: Towards Creative Machines*, Atlantis Thinking Machines. Atlantis-Springer. chapter 12, 243–266.
- Lerdahl, F., and Jackendoff, R. 1983. *A generative theory of tonal music*. Cambridge, MA: The MIT Press.
- Manurung, H. 2003. *An evolutionary algorithm approach to poetry generation*. Ph.D. Dissertation, University of Edinburgh.
- Misztal, J., and Indurkha, B. 2014. Poetry Generation System With an Emotional Personality. In *Proceedings of 5th International Conference on Computational Creativity, Ljubljana, Slovenia, ICC 2014*.
- Nakamura, C., and Onisawa, T. 2009. Music/lyrics composition system considering user's image and music genre. In *Proceedings of 2009 IEEE international Conference on Systems, Man and Cybernetics*, 1764–1769. San Antonio, TX, USA: IEEE Press.
- Netzer, Y.; Gabay, D.; Goldberg, Y.; and Elhadad, M. 2009. Gaiku: generating Haiku with word associations norms. In *Proceedings of the NAACL 2009 Workshop on Computational Approaches to Linguistic Creativity, CALC '09*, 32–39. Boulder, Colorado: ACL Press.

- Newman, D.; Lau, J. H.; Grieser, K.; and Baldwin, T. 2010. Automatic Evaluation of Topic Coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, 100–108. Los Angeles, CA, USA: ACL Press.
- Ramakrishnan A, A., and Devi, S. L. 2010. An alternate approach towards meaningful lyric generation in Tamil. In *Proceedings of NAACL HLT 2010 2nd Workshop on Computational Approaches to Linguistic Creativity*, CALC '10, 31–39. Los Angeles, CA, USA: ACL Press.
- Shen, H.-C. 2013. Lyrical Generation for a MIDI-to-Singing System. *Advanced Science Letters* 19(7):1852–1856.
- Smith, M. R.; Hintze, R. S.; and Ventura, D. 2014. Nehovah: A Neologism Creator Nomen Ipsum. In *Proceedings of 5th International Conference on Computational Creativity, Ljubljana, Slovenia*, ICCV 2014.
- Toivanen, J. M.; Järvisalo, M.; and Toivonen, H. 2013. Harnessing Constraint Programming for Poetry Composition. In *Proceedings of the 4th International Conference on Computational Creativity*, ICCV 2013, 160–167. Sydney, Australia: The University of Sydney.
- Toivanen, J. M.; Toivonen, H.; and Valitutti, A. 2013. Automatical Composition of Lyrical Songs. In *Proceedings of 4th International Conference on Computational Creativity*, ICCV 2013, 87–91. Sydney, Australia: The University of Sydney.
- Tomašič, P.; Žnidaršič, M.; and Papa, G. 2014. Implementation of a slogan generator. In *Proceedings of 5th International Conference on Computational Creativity, Ljubljana, Slovenia*, ICCV 2014, 340 – 343.
- Turney, P. D. 2001. Mining the Web for Synonyms: PMI–IR versus LSA on TOEFL. In *Proceedings of 12th European Conference on Machine Learning, ECML 2001*, volume 2167 of LNCS, 491–502. Freiburg, Germany: Springer.
- Veale, T., and Hao, Y. 2008. A Fluid Knowledge Representation for Understanding and Generating Creative Metaphors. In *Proceedings of 22nd International Conference on Computational Linguistics*, volume 1 of COLING '08, 945–952. Manchester, UK: ACL Press.