

Classificador de Poemas

Paulo A. Araújo^(*) e Nuno J. Mamede^(**)

^(*)L²F / Instituto Superior de Engenharia de Lisboa
R. Conselheiro Emídio Navarro, 1, 1949-014 Lisboa
E-mail: paraujo@isel.pt, Telefone: 218317217, Fax: 218317114

^(**)L²F / Instituto Superior Técnico
Rua Alves Redol, 9, 1000-029 Lisboa
E-mail: Nuno.Mamede@inesc-id.pt, Telefone: 213100367, Fax: 213145843

Resumo. *Descreve-se um classificador para poemas da poética portuguesa. A classificação dos poemas é realizada com base nos conceitos de estrofe, verso, sílaba e rima definidos no dicionário de termos literários [1] e no dicionário de literatura [2]. Foi implementado um conjunto de regras que tem em conta o número de versos por estrofe, o número de sílabas de cada verso e a rima utilizada.*

O protótipo desenvolvido agrega um módulo que contém o léxico, um módulo que realiza a interface com uma aplicação externa que gera transcrições fonéticas e gera divisões silábicas das palavras, e 3 módulos funcionais.

São utilizadas técnicas de processamento da língua natural para realizar a correcção ortográfica do poema, tendo também sido desenvolvidos algoritmos para classificar os poemas.

É feita uma avaliação do protótipo descrito, utilizando diferentes tipos de poemas. Por último são apresentadas algumas conclusões relativamente às potencialidades e limitações do classificador descrito.

1. Motivação

Cada vez mais se utilizam ferramentas de apoio à escrita, como por exemplo os correctores ortográficos e gramaticais. A sua utilização reflecte-se quer na qualidade final dos documentos quer nas facilidades que proporcionam. Os poemas podem ser considerados um caso mais específico de documentos em que a sua estrutura e regras de construção são mais rígidas.

Para a língua Portuguesa não foram encontradas ferramentas para tratamento de poemas. No entanto foram encontradas páginas na internet onde é possível ler e discutir poesia. Também para a língua inglesa existem este tipo de páginas na internet mas também existem outras onde é possível jogar com palavras. O “The Gardener Kit” [3] é um jogo de palavras em que o objectivo final é a construção de poemas. Este jogo corresponde a uma versão para a internet inspirada na versão original “Magnetic Poetry” de Dave Kapell [4]. Também foram encontrados sistemas mais complexos que permitem a geração automática de poesia como é o caso do “Ray Kurzweil’s Cybernetic Poet” [5] embora neste caso se utilizem modelos de linguagem para representação de poemas pré-fornecidos.

A motivação para a realização deste trabalho foi a implementação de um sistema capaz de classificar poemas e servir de incentivo para a realização, leitura e estudo da poesia. Poderá ser utilizado como uma ferramenta

didáctica de apoio ao estudo de poesia nas escolas, para poetas que realizam poesia, e que poderão ver algumas das tarefas de verificação dos poemas simplificadas e, ainda, como auxiliar de leitura em voz alta de poemas.

2. Arquitectura

A arquitectura agrega um processo central que é responsável por coordenar os vários módulos que compõem o protótipo construído. Foi inspirada na plataforma Galaxy-II [6] e é resumida na Figura 1.

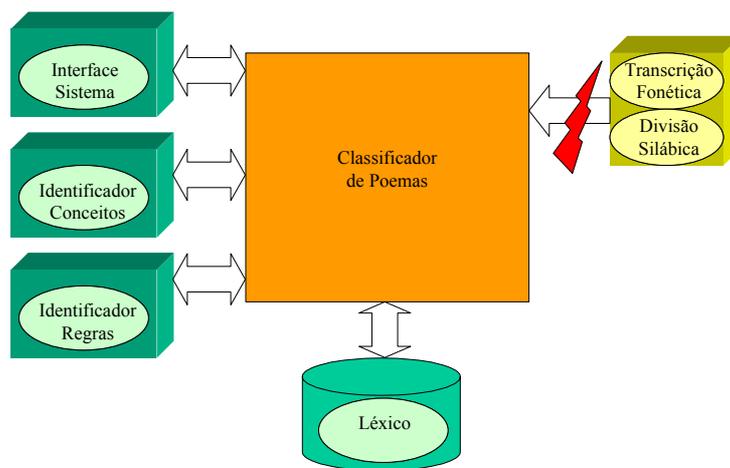


Figura 1 – Arquitectura do protótipo realizado.

O módulo que realiza a interface com a aplicação externa designa-se por *módulo de interface externa (MIE)*. A aplicação externa gera *transcrições fonéticas* e *divisões silábicas* e o seu funcionamento está baseado no sistema DIXI [7] que tem como base um sistema de regras. Este módulo tem duas funções principais que podem ser aplicadas a uma palavra ou a um conjunto de palavras. Uma função invoca os procedimentos necessários para realizar a transcrição fonética, e a outra para realizar a divisão silábica.

O módulo que realiza interface com o léxico designa-se por *módulo de interface do léxico (MIL)*. O léxico inclui as transcrições fonéticas e as divisões silábicas fornecidas pela aplicação externa e à qual se acrescenta informação adicional para facilitar o processo de classificação.

O módulo que realiza a interface com o utilizador designa-se por *módulo de interface do sistema (MIS)*. As operações de interface com o utilizador incluem a edição de poemas, guardar e ler em ficheiros os poemas editados e recepção dos comandos do utilizador para validação de vocabulário e para classificação dos poemas.

O módulo que identifica os conceitos da poética portuguesa, designa-se por *módulo identificador de conceitos (MIC)*, é responsável por reconhecer e assinalar num poema os conceitos estruturais e a rima do poema.

O módulo que atribui a classificação final do poema designa-se por *módulo identificador de regras (MIR)*. Este módulo implementa um conjunto de regras de classificação.

O processo coordenador dos módulos foi designado por *processo classificador de poemas (PCP)*.

3. Implementação

O *PCP* realiza a classificação de poemas em 4 etapas. Na primeira etapa, o *MIS* é responsável por realizar a aquisição do poema para uma estrutura interna e armazena o poema num formato que identifica as linhas do poema e destaca as palavras finais dos versos. Este formato simplifica a etapa de classificação seguinte. Pode, opcionalmente, ser realizada a verificação de vocabulário do poema, utilizando as funções do *MIL*, para garantir que as palavras que compõem o poema existem no léxico. Se tal não acontecer, as palavras em falta podem ser acrescentadas ao léxico utilizando funções do *MIE*.

Na segunda etapa o *MIC* acrescenta a informação de identificação de conceitos. Esta informação inclui a identificação das estrofes, dos versos e das rimas. Também é acrescentada a informação da transcrição fonética e divisão silábicas das palavras.

Na terceira etapa o *MIR* realiza a classificação do poema com base em regras que incluem a classificação das estrofes, a classificação dos versos e a classificação da rima. Tem ainda como base a informação de identificação de conceitos da etapa anterior.

Na quarta e última etapa, o *MIS* é responsável por apresentar o resultado final de classificação ao utilizador.

3.1. Identificação dos Conceitos de *Poema*, *Estrofe* e *Verso*

O *MIC* identifica os conceitos de *poema*, *poesia*, *estrofe* e *verso*. As definições que foram usadas para definir estes conceitos podem ser analisados a 2 níveis que correspondem ao detalhe e à objectividade da definição. Por *poema* entende-se “*um organismo verbal que contém, suscita ou segrega poesia*” [1]. Como a definição de *poema* está baseada na definição de *poesia* então para *poesia* adoptou-se a seguinte definição:

“*Se eu chamar prosa a um discurso mínimo, veículo mais económico do pensamento, e chamar, a, b, c, a atributos particulares da linguagem, inúteis mas decorativos, tais como o metro, a rima ou o ritual das imagens, toda a superfície das palavras se encaixará na dupla equação de M. Jourdain:*

$$\text{Poesia} = \text{Prosa} + a + b + c$$

$$\text{Prosa} = \text{Poesia} - a - b - c$$

Daí resulta evidentemente que a Poesia é sempre diferente da Prosa. Mas tal diferença não é de essência, é de quantidade” [1].

Pode-se concluir destas duas definições que poesia é diferente de prosa e essa diferença encontra-se nos atributos particulares da linguagem, tais como o metro e a rima. Daqui resulta o primeiro compromisso em relação à implementação destes conceitos que restringe o domínio dos poemas que podem ser analisados pois apenas se consideram poemas válidos aqueles em que existe poesia.

Os conceitos de *estrofe* e *verso* estão relacionados com a estrutura do poema. “*Por estrofe entende-se cada uma das secções que constituem um poema, ou seja cada agrupamento de versos, rimados ou não, com unidade de conteúdo e de ritmo*” [1]. “*Verso é a sucessão de sílabas ou fonemas formando unidade rítmica e melódica, correspondente a uma linha do poema. Cada verso subdivide-se ainda em sub unidades*

caracterizadas pelo agrupamento de sílabas chamado de pé na versificação greco-latina” [1]. A partir das duas definições apresentadas se conclui que um poema está organizado em secções que se designam de estrofes e as linhas do poema são designadas por versos. A Figura 2 mostra a estrutura de um poema.

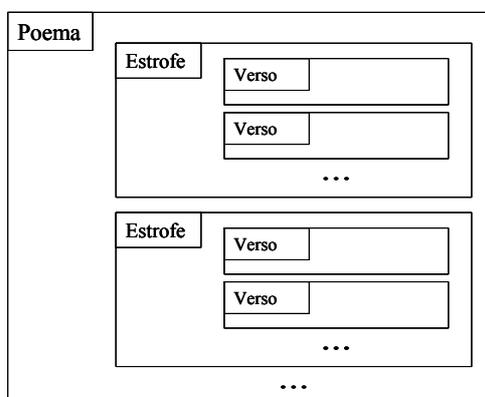


Figura 2 – Estrutura do poema

O algoritmo que implementa a identificação dos conceitos é resumido na Tabela 1. Este algoritmo está implementado no *MIC*. Começa por iniciar as variáveis *numEstrofe* e *numVerso* com 0 para depois realizar um ciclo de leitura das linhas do poema. Por cada linha lida incrementa a variável *numVerso* que indica o número de versos numa estrofe. Por cada mudança de linha em branco no texto é incrementada a variável *numEstrofe* que indica o número de estrofes. Por cada mudança de estrofe é iniciada a variável *numVerso* com 0.

Tabela 1 – Algoritmo de identificação de Conceitos

```

// Inicia variáveis
numEstrofe=0
numVerso=0
// Inicia estrutura interna
NovoPoema();
// Ciclo de identificação de conceitos
while((txtLinha=bufRead.readLine())!=null) {
  // ... pré processamento da linha
  if(txtLinha.compareTo("")!=0)
  { // caso a linha não seja nula
    numVerso=numVerso+1 // incrementa o nº verso
    Processa(txtLinha, numVerso)
  }
  else { // caso a linha seja nula
    if(numVerso>0) { // salta linhas brancas consecutivas
      // incrementa o número de estrofe
      numEstrofe=numEstrofe+1;
      ActualizaEstrofe(numEstrofe, numVerso);
      // actualiza a estrofe e reinicia o numVerso
      numVerso=0;
    }
  }
}
}

```

3.2. Identificação de Rima

A Identificação da Rima também é realizada pelo *MIC* e realiza-se em duas fases. A primeira fase é realizada na instrução “Processa(txtLinha, numVerso)” que faz parte do algoritmo da Tabela 1. Esta instrução decompõe cada linha do poema em palavras, e para a última palavra acrescenta a informação adicional necessária para a rima. A segunda fase é realizada no fim da aquisição do poema e é feita da seguinte forma. Começando pela última palavra do 1º verso do poema, atribui-se a letra *A* a essa palavra. Para todas as

palavras finais dos versos do poema que rimem com essa palavra, também se atribui a mesma letra. Procura-se a próxima palavra final que ainda não tenha letra associada e atribui-se outra letra, por exemplo *B* e repete-se sucessivamente este procedimento até todas as palavras finais terem letra atribuída.

O conceito de rima foi definido a partir da seguinte forma: “*Depara-se-nos uma rima (final) quando, em duas ou mais palavras, a última vogal acentuada, com tudo o que se lhe segue, tem idêntica sonoridade*” [1].

A Figura 3 esquematiza a definição apresentada.

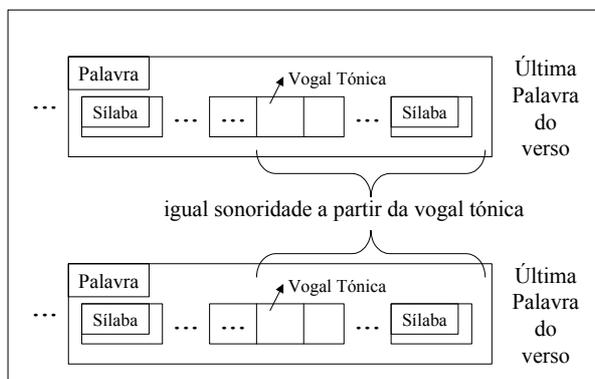


Figura 3 – Rima

A comparação sonora das palavras é realizada com base na transcrição fonética das palavras. A aplicação externa utiliza o alfabeto fonético SAM-PA para realizar as transcrições fonéticas.

A título de exemplo as palavras ‘tradução’ e ‘abalaram’ têm a transcrição fonética respectivamente de [tr6dus"6~w~] e [6b6l"ar6~w~]. Estas traduções correspondem à transcrição fonética no alfabeto fonético SAM-PA e a vogal tônica é precedida do carácter “'”. Para verificar se duas palavras rimam basta comparar as transcrições fonéticas a partir do carácter “'”.

3.3. Regras de classificação de Poemas

O módulo *MIR* realiza a classificação dos poemas com base em regras. Para isso foram implementadas um conjunto de tabelas de classificação a partir das quais se classifica os poemas. O número de versos em cada estrofe condiciona a classificação das estrofes que compõem o poema. O número total de sílabas de cada verso pode variar ao longo do poema. Consoante o número de sílabas que compõem o verso assim se obtêm diferentes tipos de verso. A escolha das palavras finais e a forma como a rima é enlaçada define o tipo de rima utilizada.

Começando pela classificação das estrofes quanto ao número de versos a Tabela 2 mostra as várias designações de classificação possíveis e a designação escolhida:

Nº de Versos	Designações possíveis	Designação Adoptada
1	monótico	monótico
2	dístico, parelha ou pareado	parelha
3	trístico ou terceto	terceto
4	tetrástico, quadra ou quarteto	quadra
5	pentástico, quinteto ou quintilha	quinteto
6	hexástico, sextilha, sexteto ou septena	sexteto
7	heptástico, sétima, septilha, septena ou hepteto	sétima

8	octástico ou oitava	oitava
9	nona, eneagésima ou novena	nona
10	decástico, década ou décima	décima
n	n versos	n versos

Devido à necessidade de utilização de apenas uma das designação na classificação de poemas, foi escolhida a mais comum. A linha n corresponde a uma generalização no caso do número de versos de uma estrofe ser maior que 10 e generaliza os restantes casos.

A próxima classificação depende do número de sílabas que existe em cada verso, e para cada caso também se existem várias designações possíveis. A Tabela 3 resume os casos possíveis.

Tabela 3 – Classificação dos versos quanto ao nº de versos

Nº de Sílabas	Designações possíveis	Designação Adoptada
1	1 sílaba	1 sílaba
2	dissílabo ou bissílabo	bissílabo
3	trissílabo, quebrado de redondilha maior, redondilho quebrado ou cola	trissílabo
4	Tetrassílabos	tetrassílabos
5	pentassílabo ou redondilha menor	pentassílabo
6	hexassílabo, heróico quebrado ou heróico menor	hexassílabo
7	heptassílabo ou redondilha maior	heptassílabo
8	Octossílabo	octossílabo
9	eneassílabo, verso de gregório de matos	eneassílabo
10	decassílabo, heróico, sáfico ou provençal	decassílabo
11	hendecassílabo ou verso de arte maior	hendecassílabo
12	alexandrino	alexandrino
m	m sílabas	m sílabas
20	vintissílabos	vintissílabos
n	n sílabas	n sílabas

Também neste caso se adoptou a designação mais comum para ser utilizada. As linhas m e n correspondem à generalização para os casos que em que o número de versos não se encontra na tabela e que correspondem aos valores entre 13 e 19 e aos maiores que 20.

A última palavra de cada verso determina o tipo de rima utilizado. Existem três tipos possíveis de classificação que são resumidas na Tabela 4:

Tabela 4 – Classificação quanto ao tipo de rima

Designações possíveis	Designação adoptada
oxítona ou aguda	aguda
paroxítona ou grave	grave
proparoxítona ou esdrúxula	esdrúxula

Também aqui o critério de escolha para a designação a adoptar foi a que é mais usual.

Para classificar a posição relativa dos versos e a forma como estão enlaçados foi realizada a Tabela 5 com as várias hipóteses possíveis:

Tabela 5 – Classificação quanto à posição relativa dos versos que enlaça

Enlace da Rima	Designação
Quando os versos que rimam se encontram juntos e aos pares (ABB CDD EFF)	emparelhadas
Quando entre dois versos que rimam se encontra outro de diferente rima (ABCB ou ABAB)	cruzadas
Quando entre dois versos que rimam se encontram dois versos de diferente rima (ABBA ou ABCA)	abraçadas
Quando entre dois versos que rimam se encontram três ou mais	interpolada

versos de diferente rima (ABBBA ou ABCDA)	
Quando rimam mais de dois versos seguidos (AAA)	seguida
Quando está sujeito a uma só rima que pode também ser cruzada (ABCB)	monórrimos

4. Exemplo de Classificação

4.1. Tipos de Poemas

Assume-se como diferentes tipos de poemas, poemas que apresentem diferentes hipóteses de classificação nas tabelas apresentadas.

Para exemplificar diferentes tipos de poemas, foram utilizados poemas com diferentes proveniências, sendo o primeiro exemplo 2 quadras populares de António Aleixo, o segundo exemplo 1 estrofe dos lusíadas e o terceiro exemplo duas estrofes realizadas por crianças em idade escolar.

4.2. Resultados de Classificação

Na Tabela 6 apresentam-se os exemplos de classificação dos poemas.

Tabela 6 – Classificação de Poemas	
Poemas	Classificações
Eu não tenho vistas largas, Nem grande sabedoria, Mas dão-me as horas amargas Lições de Filosofia.	Classificação do Poema Resumo linhas: 9 ; versos: 8 ; estrofes: 2 sílabas: [8,8,8,8,0,8,9,8,8] rimas: [A,B,A,B, ,C,D,C,D] Classificação das estrofes 1ª estrofe - quadra [4 versos] - rima cruzada 2ª estrofe - quadra [4 versos] - rima cruzada
Há luta por mil doutrinas. Se querem que o mundo ande, Façam das mil pequeninas Uma só doutrina grande.	Classificação do Poema Resumo linhas: 8 ; versos: 8 ; estrofes: 1 sílabas: [12,12,12,13,12,13,13,11] rimas: [A,B,A,B,A,B,C,C] Classificação por estrofes 1ª estrofe - oitava [8 versos] - rima cruzada, rima emparelhada.
As armas e os barões assinalados Que da ocidental praia lusitana Por mares nunca de antes navegados Passaram ainda além da Taprobana, Em perigos e guerras esforçados Mais do que prometia a força humana, E entre gente remota edificaram Novo Reino, que tanto sublimaram;	Classificação do Poema Resumo linhas: 8 ; versos: 7 ; estrofes: 2 sílabas: [9,8,8,8,0,4,5,4] rimas: [A,B,C,B, ,C,C,C] Classificação por estrofes 1ª estrofe - quadra [4 versos] - rima cruzada 2ª estrofe - terceto [3 versos] - rima seguida
menina que leva a vida sentadinha a escrever, faça favor de ensinar, eu também quero aprender.	
Brincar, brincar és para brincar e alegrar	

O resultado de classificação divide-se em duas partes. Na primeira parte é feito um resumo do poema onde se pode verificar que o número de linhas total do poema, o número de versos e o número de estrofes. O resumo inclui também a informação do número de sílabas em cada verso do poema e ainda o enlace da rima. Na segunda parte é feita a classificação das várias estrofes que compõem o poema.

A informação de classificação das estrofes inclui a classificação quanto ao número de versos e quanto à rima utilizada. Podem existir várias hipóteses de rima na mesma estrofe.

Existe uma opção de classificação que permite aumentar o detalhe de classificação. A Tabela 7 mostra o resultado que se obteria para a primeira quadra do primeiro exemplo apresentado.

A informação de detalhe mostra a classificação de cada verso quanto ao número de sílabas e quanto ao tipo da última palavra. Inclui ainda o enlace da rima e o tipo de rima.

1ª estrofe - quadra [4 versos] - rima cruzada
1º Verso - octossílabo [8 sílabas] - grave - A
2º Verso - octossílabo [8 sílabas] - grave - B
3º Verso - octossílabo [8 sílabas] - grave - A
4º Verso - octossílabo [8 sílabas] - grave - B

5. Conclusões

Em relação aos objectivos iniciais, o protótipo realiza a classificação de diferentes tipos de poema sem ser necessário à partida fornecer exemplos de poemas.

Utiliza-se o mesmo processo de classificação para classificar os diferentes tipos de poemas e com diferentes proveniências.

À medida que se foram fazendo experiências de classificação com diferentes tipos de poesia, apenas no início se notou mais a falta de vocabulário no léxico que, à medida que foi sendo acrescentada fez diminuir as palavras em falta.

Para poemas mais extensos nota-se que a performance do sistema diminui devido ao aumento do número de palavras que obrigam a um maior processamento.

6. Referências

- [1] Moisés, M., *Dicionário de Termos Literários*, São Paulo: Brasil, Editora Coltrix, 1974.
- [2] Coelho, J., *Dicionário de Literatura*, Porto-Figueirinhas: Portugal, Editora do Minho 3ª Edição, 1987
- [3] *The Gardener Kit*, Magnetic Poetry Inc., 2001
<http://www.magneticpoetry.com/magnet/index.html>
- [4] *The Magnetic Poetry Story*, Magnetic Poetry Inc., 2001
<http://www.magneticpoetry.com/story.html>
- [5] Ray Kurzweil, *Ray Kurzweil's Cybernetic Poet*, Kurzweil CyberArt Technologies, 1999
http://www.kurzweilcyberart.com/poetry/rkcp_overview.php3
- [6] Seneff, S., Hurley, E., Lau, R., Pao, C., Schmid, P. e Zue, V., "Galaxy-II: A Reference Architecture for Conversational System Development" in Proc. ICSLP '98, Sydney, Australia, 30 Nov.-4 Dec. 1998, 3:931-934.
<http://www.sls.lcs.mit.edu/sls/publications/1998/icslp98-galaxy.pdf>
- [7] Oliveira L., *DIXI - Sistema de síntese de fala a partir de texto*.