

O USO DE PROVÉRBIOS NO ENSINO DE PORTUGUÊS

Sónia Reis, Universidade do Algarve, *Portugal*

Jorge Baptista, Universidade do Algarve e INESC-ID Lisboa, *Portugal*

Resumo

Os provérbios, enquanto elementos linguísticos marcados pelo seu uso generalizado, são estruturas muito ricas, quer do ponto de vista cultural, quer do ponto de vista linguístico, daí o seu potencial como um importante contributo para o ensino das línguas. É neste sentido que se justifica a sua inclusão nos conteúdos programáticos de Português, em articulação com as respetivas Metas Curriculares. Apesar de os provérbios apenas constarem explicitamente dos conteúdos programáticos do ensino básico de português do 5º e 6º anos de escolaridade, os mesmos surgem disseminados nos manuais escolares até ao 12º ano de escolaridade. Neste trabalho, pretendemos, num primeiro momento, verificar em que alturas do percurso escolar são os provérbios referidos tanto nos textos ordenadores como nos manuais escolares, bem como determinar os fatores didáticos que promovem a sua utilização. Num segundo momento, pretendemos determinar quais os provérbios (e suas variantes) mais difundidos nos manuais escolares, com recurso a ferramentas do processamento de linguagem natural. Os resultados permitirão extrair índices de frequência e repartição dos provérbios nos manuais escolares e construir uma imagem mais clara da forma como são utilizados no contexto de ensino de língua e o potencial impacto que tal representação poderá ter na respetiva disponibilidade lexical.

Palavras-chave: Disponibilidade lexical, ensino de língua, manuais escolares, Processamento da Linguagem Natural (PLN), provérbios portugueses

1. Introdução

Os provérbios apresentam uma grande diversidade das suas estruturas e, sendo muitas vezes frases curtas e concisas, têm servido como ferramenta para o estudo e o ensino da língua nas suas diversas componentes - morfologia, fonologia, sintaxe, semântica, pragmática, entre outros.

Dado a sua riqueza linguística e cultural, os provérbios surgem disseminados nos manuais escolares de Português dos diferentes níveis de escolaridade e servindo os mais variados propósitos. Perante isto, pretendemos verificar em que níveis de ensino-aprendizagem são os provérbios referidos nos manuais escolares e os fatores didáticos que promovem a sua utilização, bem como determinar quais os provérbios (e suas variantes) mais difundidos nos mesmos manuais.

2. Enquadramento

Os provérbios são utilizados no ensino do Português, quer enquanto língua materna, quer enquanto língua não materna. Surgem ainda no *Programa de Português L2 [= Língua Segunda] para Alunos Surdos* (Baptista et al. 2011).

Nesta secção, faremos uma breve análise de alguns dos documentos ordenadores de ensino de Português com o objetivo de caracterizar a forma como estes documentos enquadram os provérbios no processo de ensino/aprendizagem da língua.

2.1. Português língua materna

No que respeita ao ensino de Português enquanto língua materna, os provérbios surgem explicitamente como conteúdo do 2º ciclo de estudos¹, integrados no *Programa e Metas Curriculares do Português do Ensino Básico* (Buescu et al. 2015).

As *Metas Curriculares de Português de Ensino Básico* têm por principal objetivo melhorar a qualidade do ensino e da aprendizagem nos nove anos do ensino básico, definindo os marcos a atingir pelos alunos no final de cada ano letivo, os quais são avaliados por meio de descritores de desempenho. No *Programa*, são definidos os conteúdos por ano de escolaridade bem como a organização sequencial e hierárquica dos mesmos.

No 1º e 2º ciclos, os conteúdos estão organizados em quatro domínios - Oralidade, Leitura e Escrita, Educação Literária, Gramática – e, no 3º ciclo, em cinco domínios, tendo-se autonomizado os domínios da Leitura e da Escrita (*idem*: 3). Os provérbios estão incluídos no domínio da Leitura e Escrita do 2º ciclo de estudos. Neste sentido, os alunos deverão “[f]azer inferências a partir da informação prévia ou contida no texto - Identificar, pelo contexto, o sentido de palavras, expressões ou fraseologias desconhecidas, incluindo provérbios e expressões idiomáticas” (*idem*: 70).

Apesar de os provérbios apenas constarem explicitamente no domínio da Leitura e Escrita, do 5º e 6º ano, estes surgem disseminados nos manuais escolares até ao 12º ano de escolaridade. Tendo isto em conta, pretendemos verificar sistematicamente em que anos de escolaridade é que os manuais escolares fazem referência a provérbios, quando é que estes são utilizados e, numa outra fase deste trabalho, pretendemos determinar ainda quais os provérbios que aparecem nos mesmos manuais bem como a sua frequência.

2.2. Português língua não materna

De modo a harmonizar a aprendizagem e o ensino das línguas vivas na Europa, foi criado o *Quadro Europeu Comum de Referência para as Línguas* (QECR) (Conselho da Europa, 2001), juntamente com o *Portfólio Europeu das Línguas* (PEL) (Schneider *et al.* 1999). A edição portuguesa do QECR, adaptado à realidade portuguesa, foi elaborado pelo Conselho da Europa, no âmbito do projeto *Políticas Linguísticas para uma Europa Multilíngue e Multicultural*.

Segundo o QECR, as competências comunicativas em língua incluem três componentes: as competências linguísticas, as competências sociolinguísticas e as competências pragmáticas. A competência lexical, parte integrante das competências linguísticas, diz respeito ao conhecimento e utilização do vocabulário de uma língua e abrange elementos gramaticais e elementos lexicais, incluindo provérbios e expressões idiomáticas. A competência sociolinguística, enquanto conhecimento e capacidade para lidar com a dimensão social do uso da língua, também se faz valer dos provérbios, uma vez que são “fórmulas fixas, que exprimem e reforçam as atitudes correntes, [e] contribuem significativamente para a cultura popular” (Conselho da Europa, 2001: 170).

Na realidade portuguesa, o universo escolar é constituído por uma população heterogénea do ponto de vista cultural e linguístico, daí o sistema educativo procurar responder às necessidades desta comunidade escolar. Desta forma, o Ministério da Educação produziu em julho de 2005 um documento orientador – *Português Língua Não Materna no Currículo Nacional* (Leiria *et al.* 2008), que estabelece os princípios e as linhas orientadoras para a integração dos alunos dos ensinos básico, secundário e recorrente. Neste documento não há uma referência explícita a provérbios, mas os mesmos poderão estar englobados nos conceitos “expressão idiomática” (*idem*: 9), “padrões frásicos básicos com expressões memorizadas” (*idem*: 34) ou “expressões e fórmulas coloquiais” (*idem*: *ibidem*).

Num trabalho futuro, pretendemos verificar sistematicamente nos manuais usuais de PLE em que momentos do percurso escolar são referidos os provérbios, bem como determinar os fatores didáticos que promovem a sua utilização. Este estudo vai incidir exclusivamente sobre o lugar dos provérbios no ensino de Português enquanto língua materna, que tratamos nas secções 4 e 5.

2.2.3. Português língua segunda para alunos surdos

Os provérbios também são explicitamente referidos no *Programa de Português L2 para Alunos Surdos* (Baptista *et al.* 2011), fazendo parte dos conteúdos, quer do 1º, quer do 2º ciclo de estudos. Das várias tipologias textuais exploradas nestes ciclos de estudos, o provérbio é um dos tipos textuais abordados: Efetivamente, o programa do 1º ciclo pretende apresentar: “textos que evidenciam a escolha das palavras e/ou das imagens: adivinha, poema, poema visual, provérbio, acróstico, slogan, mensagem publicitária” (*idem*: 52). Também no 2º ciclo, entre os vários materiais propostos, se apresentam “[t]extos para usar como ferramenta: dicionários (visual, monolíngue, de sinónimos, de antónimos, de provérbios, de LGP, etc.), mapa, atlas, catálogo, glossário, enciclopédia ilustrada, tradutor em linha de português - LGP, gramática, prontuário.” (*idem*: 75)

3. Trabalhos relacionados

De entre os trabalhos que dizem respeito ao uso dos provérbios no ensino de língua (Martins (2010); Ferreira (2011); Duarte (2013); Ferreira & Vieira (2013); Pereira (2015), é possível destacar os que, sucintamente, apresentamos em seguida.

Martins (2010) apresenta uma unidade didática que, simultaneamente, “prevê e estimula não só a análise, mas também o uso de provérbios por alunos de Português Língua Estrangeira”. Esta unidade é direcionada para alunos de nível avançado (nomeadamente os níveis C1 e C2, de acordo com o Quadro Europeu Comum de Referência para as Línguas) e tem como principais objetivos: “a) desenvolver conteúdos linguístico-culturais avançados; b) apresentar códigos, valores e características próprias da sociedade portuguesa; c) motivar e estimular a empatia do aprendente > comunidade-alvo e d) promover a discussão intercultural” (*idem*: 97).

Nesta unidade didática, os provérbios são utilizados em diferentes tipos de exercícios:

- i. para introduzir o tópico, o autor coloca a questão aberta: “Conhece algum provérbio em Português?”;
- ii. é pedido ao aluno que analise um excerto de um texto literário que apresenta um provérbio, e que identifique esse provérbio e o “discuta com os colegas e o professor”;
- iii. é apresentado um exercício para fazer corresponder as duas partes de um provérbio que aparecem em dois quadros e, após esta fase, identificar “o sentimento, valor ou

- ideia em cada um dos provérbios”;
- iv. a partir deste exercício, o aluno deverá selecionar três provérbios e apresentar outros três com sentido equivalente ou oposto;
 - v. num outro exercício pede-se ao aluno que identifique os provérbios cujo conteúdo está representado sob a forma de uma imagem;
 - vi. é apresentado um exercício de palavras-cruzadas que tem alguns provérbios e cujo objetivo é completar a palavra em falta nesses provérbios;
 - vii. partindo da afirmação “Língua e Cultura são conceitos indissociáveis”, pede-se ao aluno que, num texto com o mínimo de 50 linhas, explique “Em que medida os provérbios [em geral] estão relacionados com esta afirmação?”

Os exercícios foram desenvolvidos em torno da obra literária *Ensaio sobre a cegueira* de José Saramago (1995)² e todos os provérbios utilizados ou constam dessa obra ou são relativos ao tema da cegueira. Talvez seja essa a razão de os provérbios usados nos exercícios desta unidade didática não serem dos mais usuais, ou as suas variantes não serem as mais representativas, apesar de todos constarem das nossas listagens de provérbios. De seguida, apresentamos alguns desses provérbios:

*Candeia que vai adiante, alumia duas vezes.
Fui a casa da vizinha, envergonhei-me, voltei para a minha, remediei-me.
Morrendo o bicho, acaba-se a peçonha.*

Apesar de esta unidade didática ter sido elaborada para alunos de um nível de proficiência avançado, pensamos, por um lado, que deveriam ter sido selecionados provérbios mais usuais e mais facilmente interpretáveis a fim de poderem ser reutilizados corretamente pelos alunos num contexto apropriado. Por outro lado, os provérbios aqui utilizados dizem respeito à obra literária em estudo e não correspondem a uma amostra de provérbios especificamente selecionada para se adequarem melhor ao público-alvo.

Ainda que a unidade didática apresentada seja apenas uma proposta de trabalho, seria interessante avaliar a sua eficiência pedagógica por forma a verificar se os objetivos propostos poderiam ser efetivamente alcançados seguindo a metodologia proposta ou que problemas se teriam levantado na sua aplicação. Infelizmente, o autor não apresenta neste artigo a avaliação desta experiência didática, não tendo sequer indicado uma metodologia para a sua avaliação.

Vejamos agora outro trabalho que trata desta temática. Pereira (2015) propôs 20 atividades dirigidas a alunos de português língua materna (PLM) e de português língua não materna (PLNM), de vários níveis, para o desenvolvimento de competências linguístico-fraseológicas. Para tal, o autor recorreu a duas abordagens diferentes, que explicitaremos de seguida.

Relativamente aos alunos de PLM, foi proposto um conjunto de 3 atividades de diagnóstico, a serem desenvolvidas em sala de aula. Para cada atividade, foi preparado um enunciado, onde o aluno regista as suas respostas e um documento correspondente (“Roteiro do Aplicador”) onde se encontram os objetivos de cada exercício, bem como o modo da sua aplicação e de registo/análise de dados. Os alunos foram divididos em 3 grupos de acordo com o nível de escolaridade em que se encontram: 2º ciclo, 3º ciclo ou secundário/cursos de Educação e Formação de Adultos (EFA). Cada aluno realizou três atividades: A primeira atividade, do domínio da compreensão do oral/expressão escrita, consistiu na *audição* e no *reconto* por escrito de um texto que continha 20 unidades fraseológicas (UF). O autor tinha como objetivo neste exercício testar a capacidade do aluno em compreender e reproduzir as UF contidas no texto. Para tal, o autor criou, então, uma grelha de registo dos dados, onde são tidos em consideração aspetos como a ausência da UF no *reconto* do texto ou o uso de uma alternativa para a mesma. A partir dos dados recolhidos, foi medida a frequência de uso para cada uma das UF, a percentagem de usos corretos, a percentagem de usos com um “desvio ligeiro” e, ainda, dos usos com um “desvio acentuado” ou “muito acentuado”³.

A segunda atividade, do domínio da escrita, teve por objetivo determinar se uma dada expressão foi compreendida e utilizada no sentido literal ou no sentido idiomático. Para tal, o aluno deveria construir uma frase da qual fizesse parte essa mesma expressão e que não deixasse margem para dúvidas relativamente ao seu sentido.

A terceira atividade, do domínio da leitura, consistiu em apresentar ao aluno uma expressão em contexto e pedir-lhe que escolhesse de entre três hipóteses a que melhor correspondia ao seu significado. Nada se diz como foram construídas estas hipóteses.

Para os alunos de PLNM, o autor adotou uma metodologia diferente. Foram utilizados os textos disponíveis *on-line* resultantes do projeto “Recolha de dados de aprendizagem de português língua estrangeira”⁴, coordenado por Isabel Leiria. A amostra consiste em 470 textos produzidos por alunos de diferentes níveis de aprendizagem (de A1 a C2), falantes de

28 línguas maternas diferentes e a frequentar à data cursos de PLE. Os textos dizem respeito a 3 áreas temáticas contempladas no projeto *Português Fundamental*: “O indivíduo”, “A sociedade”, “O meio ambiente”. O autor optou por selecionar as produções escritas que resultaram do estímulo 1.1A, pertencente à temática “O indivíduo”. Esta escolha deveu-se às seguintes razões: “(1) estímulo/tema que permite a identificação do aluno (sexo, idade, nacionalidade...) e a produção de um texto de interesse alargado; (2) nº de textos muito superior ao despoletado por qualquer outro estímulo: 141 [...], exatamente igual ao nº de informantes da nossa amostra; (3) informantes de todos os níveis de aprendizagem de PLE”. Após esta fase, o autor selecionou os 42 excertos dos textos que considerou terem um maior “interesse fraseológico” tendo como objetivo “identificar e classificar as UF (idiomáticas ou não)”, com base na classe gramatical a que pertencem. Não são explicitados, em alguma parte deste trabalho, quais os critérios seguidos pelo autor para determinar o “interesse fraseológico” dos excertos utilizados. De qualquer forma, destes textos, o autor contabilizou: “80 UF diferentes, sendo que 41 são verbais, 16 adverbiais, 7 nominais, 7 conjuncionais, 6 preposicionais e 3 adjetivais”. Verificou, também, que a distribuição da UF pelos 3 níveis de aprendizagem não é equitativa em termos de frequência de uso. O autor constatou ainda que nenhum dos alunos utilizou qualquer provérbio e que cerca de 30% das UF utilizadas têm um sentido idiomático, tendo sido empregues essencialmente pelos alunos de um nível mais avançado de PLE (de B1 a C2). O autor mediu também a qualidade de utilização das UF tendo em conta os “desvios” a que foram sujeitas, e teve igualmente em consideração os informantes que cometeram esses desvios. O próprio autor refere que estes dados poderão ser “porventura pouco rigorosos” mas que terão permitido “extrair” pistas sobre a competência fraseológica dos alunos. Nesse sentido, o autor sumariza as suas observações dizendo que:

“a conclusão mais importante é, no nosso entender, a evidência de que também em contextos de ensino-aprendizagem de PLNM os utilizadores se socorrem natural e espontaneamente das UF, mesmo das idiomáticas. [...] A multiplicidade e o grau de correção é que variam em função de determinados fatores (como nível de proficiência em PLE, LM e conhecimento de outras línguas), o que nos conduz à certeza de que é preciso criar condições para alargar e aperfeiçoar as competências fraseológicas dos aprendentes.” (Pereira, 2015: 297)

Além das três atividades de diagnóstico referidas, o autor propôs ainda um conjunto de 20 atividades dirigidas a alunos de PLM e de PLNM. Segundo o mesmo autor, deve haver

“um contacto natural e progressivo com as EI [expressões idiomáticas] desde os primeiros níveis de aprendizagem, para que o utilizador de PLM e PLNM as apreenda como elementos intrínsecos da língua”. (*idem*: vii)

e acrescenta ainda que

“nem em contextos de PLM a competência fraseológica é um dado adquirido: apresenta estádios de desenvolvimento deficitários, sendo necessário complementar atempadamente a aquisição natural de UF com um ensino promotor de competências comunicativo-fraseológicas.” (*idem: ibidem*)

Das 20 atividades propostas pelo autor, 14 fazem referência a provérbios. Verificámos que, em algumas destas atividades, é pedido ao aluno que comente um determinado provérbio, oralmente ou por escrito, ou que redija um texto no qual utilize expressões fixas ou provérbios partindo de diferentes estímulos. A tipologia textual bem como a dimensão do texto que o aluno irá redigir varia de atividade para atividade, podendo tratar-se de um texto expositivo-argumentativo, um texto de opinião, um diálogo, entre outros.

Efetivamente, algumas destas propostas de atividades poderão ser estimulantes para os alunos e poderão, eventualmente, contribuir para um melhor conhecimento deste tipo de expressões. No entanto, o autor não aplicou nem avaliou a eficiência destes materiais em situações reais. Também não explicitou em que medida é que estas atividades diferem das que tinha encontrado aquando da avaliação dos materiais didáticos existentes. Gostaríamos ainda de ter visto explicitados os critérios seguidos para a seleção destes exercícios/atividades, bem como para a escolha do material fraseológico em cada exercício.

Na secção seguinte, justamente, iremos apresentar a metodologia adotada para determinar o papel que os provérbios desempenham em manuais escolares, a sua diversidade e frequência de utilização como objeto de ensino-aprendizagem.

4. Metodologia

Para podermos determinar a distribuição de provérbios nos manuais escolares, fazendo-o de forma sistemática e tanto quanto possível automática (ou pelo menos não exclusivamente manual), é necessário desenvolver ferramentas que nos permitam identificar a sua ocorrência nos textos.

Os provérbios surgem em diferentes tipos de textos, tendo diferentes funções no discurso em que estão inseridos. Dado serem estruturas que, para além de apresentarem variação léxico-sintática (Chacoto, 1994), também permitem a reutilização dos seus elementos lexicais mais característicos, a sua identificação automática em textos torna-se um verdadeiro desafio. Por outro lado, este tipo de estruturas é muito rico, quer do ponto de vista linguístico quer do ponto de vista cultural, daí o seu potencial como um importante contributo para o ensino das línguas. Desta forma, a sua identificação automática em textos poderá ser uma mais-valia pois os resultados permitirão extrair índices de frequência e repartição dos provérbios nos textos, bem como construir uma imagem mais clara da forma como são utilizados no contexto de ensino de língua.

Para a identificação automática de provérbios e variantes em textos, está já a ser construída uma ferramenta (Reis & Baptista, 2016a), usando técnicas de máquinas de estados finitos (Paumier, 2016) com base numa recolha de mais 114 mil provérbios digitalizados a partir de quatro coletâneas de provérbios (Reis & Baptista 2016b). Após esta fase, foi necessário operar várias transformações nas listagens iniciais de forma a poder uniformizar as listas de provérbios. Estas transformações consistiram essencialmente em passar todas as palavras para minúsculas, remover todos os sinais de pontuação e remover todos os espaços em branco ou outros erros resultantes da digitalização das coletâneas. De seguida, optou-se por uma estratégia de simplificação textual removendo-se todas as palavras gramaticais que considerámos não serem necessárias para a identificação dos provérbios. Após a remoção destas *stopwords*⁶ determinou-se uma lista-chave com cerca de 52.000 expressões diferentes. Este método permitiu agrupar algumas variantes de provérbios que continham as mesmas sequências de palavras-chave. De seguida, com base nas sequências que surgiram repetidas mais vezes e progredindo por ordem decrescente de frequência, foram construídos transdutores para poder identificar automaticamente em textos todos os provérbios e as respetivas variantes que contivessem as mesmas palavras-chave. Estes transdutores dão conta de diversas variações morfossintáticas que os provérbios podem sofrer e também permitem a inserção de palavras ou de sinais de pontuação entre as palavras-chave do provérbio. Estas opções permitem, ainda, que sejam encontrados provérbios que possam ter sofrido alguns processos de reutilização e subversão estilística.

Sendo esta uma tarefa morosa e que ainda se encontra em curso, recorreu-se à lista de sequências de palavras-chave associadas aos provérbios já recolhidos de diversas coletâneas

para construir um conjunto de ferramentas que permitissem retirar do *corpus* as frases em que essas sequências ocorrem. Estas ferramentas foram de dois tipos

- i. um conjunto de expressões regulares produzido automaticamente e aplicado ao *corpus* por um comando *grep*;
- ii. um conjunto de transdutores com expressões introdutórias de provérbios e termos equivalentes.

Estes dois métodos são apresentados nas secções seguintes.

4.1. *Corpus* de manuais da Porto Editora

Para a identificação automática de provérbios em textos, utilizámos 38 manuais de Português (2,9 milhões de palavras), editados pela Porto Editora, do 5º ao 12º anos de escolaridade. Os textos utilizados dizem respeito tanto a manuais escolares, como a cadernos de atividades, cadernos de exercícios e seletas de textos. Trata-se de um *corpus* de conveniência, uma vez que é o único *corpus* que tínhamos disponível em suporte digital. Na figura seguinte, apresentamos a distribuição dos manuais por ano de escolaridade.

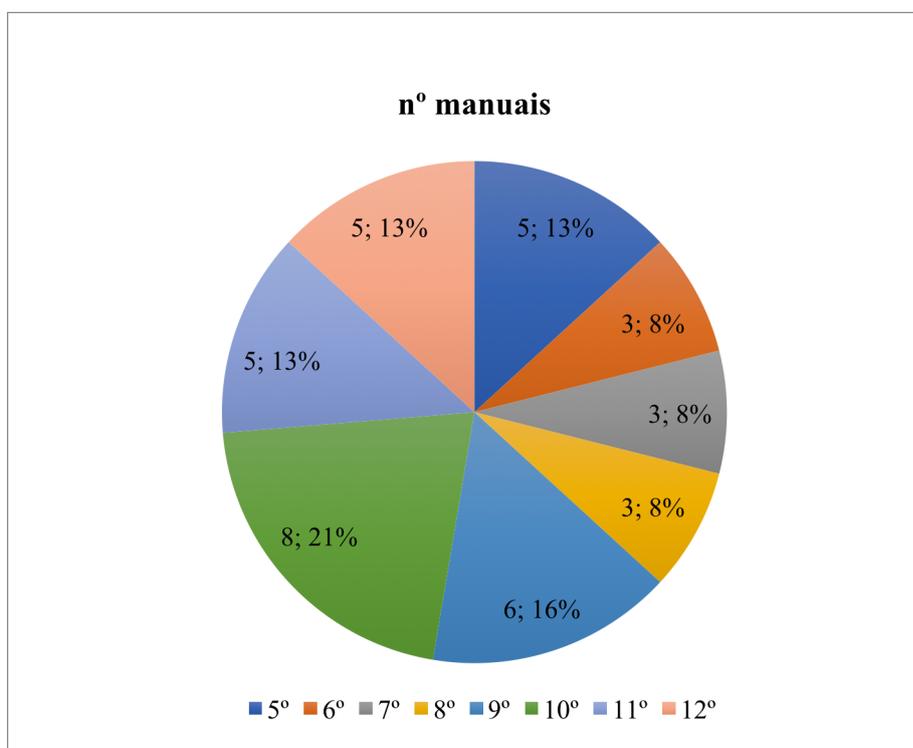


Figura 1, Distribuição dos manuais por ano de escolaridade

Como se pode ver, esta distribuição não é uniforme, havendo um número muito superior à média (4,75) no 5º e nos 9º a 12º anos, enquanto os 6º a 8º anos se encontram sub-representados. Numa primeira análise, verificou-se que apenas 4 dos 38 manuais faziam uma referência explícita a provérbios no índice.

4.2. Pesquisa de provérbios por sequência de palavras-chave associadas a provérbios (*scripts*)

Foram construídos automaticamente comandos *grep* com expressões regulares para a pesquisa de sequências de 2 a 5 palavras-chave, com uma janela de 0 a 3 palavras intervenientes e uma lista de separadores autorizados⁵. Neste trabalho, tratámos apenas os resultados das pesquisas de 3 a 5 palavras-chave, dado o elevado número de falsos positivos referentes à pesquisa com 2 palavras-chave. Como resultado, foram obtidas automaticamente 192 linhas de texto com as sequências encontradas (123 referentes a 3 palavras-chave, 45 a 4 palavras-chave e 24 a 5 palavras-chave), que após validação manual se verificou corresponderem efetivamente a 79 provérbios. De seguida, apresentamos dois exemplos das linhas de texto extraídas com os provérbios, uma com 3 palavras-chave e outra com 5 palavras-chave.

e.g. pesquisa com 3 palavras-chave

boca-morre-peixe

Pela boca morre o peixe ... e se denuncia o delinquente. Os especialistas em padrões de linguagem individual são capazes de identificar o autor de um telefonema terrorista ou de um bilhete de resgate.

e.g. pesquisa com 5 palavras-chave

#####imagem-vale-mais-mil-palavras#####

O estudante australiano sabe que jamais competirá em qualidade com as câmaras digitais mais caras, mas defende que o seu conceito poderá também ser uma alternativa aos telemóveis com câmara. Calvey vai mais longe e diz que há tecnologia para transformar a sua ideia numa proposta comercial. Afinal, não é **uma imagem que vale mais do que mil palavras?** Ou, como diria o meu Tio Olavo: **“Uma imagem vale mais que mil palavras”**. Mas o curioso é que é preciso escrever sete palavras para compor esta imagem. Qual será a opinião do autor sobre a ideia feita de que uma imagem vale mais do que mil palavras? Justifica, com base no texto.

Após esta fase foi necessário determinar o manual a que cada uma destas entradas pertencia, bem como identificar o tipo de exercício ou o tipo de texto referentes à sua utilização.

Dadas as limitações deste método, em particular, o facto de não permitir pesquisar palavras pelos seus lemas, esta pesquisa não consegue capturar certas variações, sobretudo da flexão verbal, pelo que foi necessário complementá-la utilizando outra estratégia, que apresentamos na secção seguinte.

4.3. Expressões introdutórias e lista de termos associados

Este segundo método consistiu na construção de uma gramática local usando uma ferramenta baseada em máquinas de estados finitos (Paumier, 2016) com base em expressões introdutórias: *lá dizia a minha avó, como se costuma dizer, lá diz o povo*, e também com base em termos: *provérbio, rifão, adágio, aforismo, axioma, ditado/N, parémia, prolóquio, anexim*, etc., admitindo variação morfossintática (Nome/Adjetivo: gênero-número, Verbo: pessoa-número-tempo-modo) e certas inserções.

Como resultado desta pesquisa obtivemos 239 ocorrências. Algumas destas são meras referência a estes termos, não estando associadas diretamente a provérbios presentes nos textos. Apresentamos, abaixo, um excerto do resultado da concordância onde se podem encontrar alguns desses termos ou expressões introdutórias.

conta que ele convoca o **aforismo** popular «Uma no cravo, outra na ferradura». Descreva as personagens rrigem os costumes", ou, **como diz o povo**, a rir se dizem certas verdades). - A IRONIA E O EUFEMISMO Deus é brasileiro; este **ditado**, muito conhecido e repetido no Brasil, deu o nome a um filme realiza -2005 1. Referência a um **ditado popular** que diz que Deus é brasileiro; este ditado, muito conhecido s morfologicamente. 4. O **ditado** "Quem vê caras não vê corações" pode relacionar-se com o argumento d nge do coração, **diz lá o ditado**. Ora, queira Deus que não seja por minga de saúde; e, se é, di-lo pa Teyssier) - citamos-te o **ditado popular** que diz "de sábio e de louco todos temos um pouco". E a este e a lei deles é o velho **ditado** "quem não trabuca não manduca". Não pensem que é fácil a vida de um com uma referência a um **provérbio**: "Quem conta um conto acrescenta um ponto". Sabes o que é um prov o." [verso 1] 1.1. Há um **provérbio que diz**: "O que é bom acaba depressa." Explica o sentido do prime nhos de atenção. **Reza um provérbio que** devemos dar tempo ao tempo, mas a verdade é que, dê por onde

Figura 2, Excerto de concordância de pesquisa de expressões introdutórias e termos associados

Este método apresenta, porém, algumas desvantagens. Depois de encontradas as expressões introdutórias associadas, ainda é necessário procurar os provérbios no contexto, o que mobiliza diversas competências (lexicais, culturais), implica reconhecer a estrutura do provérbio: rima, métrica, etc., ou reconhecer marcas (orto-/tipo-)gráficas distintivas: aspas, itálico, negrito, entre outras. Para melhor demonstrar as dificuldades que este tipo de tarefa poderá representar, apresentamos, abaixo, um exercício que consta de um dos manuais escolares em análise e que foi encontrado através deste método de pesquisa.

Velhos Provérbios

Um Velho Provérbio foi ver uma partida de futebol chamou à parte um jogador e sussurrou-lhe ao ouvido **Quem age sozinho age por três** O futebolista experimentou jogar à bola sozinho mas não tinha graça nenhuma nunca podia vencer e por isso voltou à equipa O Velho Provérbio desapontado adoeceu e tiveram que tirar-lhe as amígdalas Três Velhos Provérbios encontraram-se e mal abriram a boca começaram logo a desconversar **Obra começada meio acabada** disse o primeiro De maneira nenhuma disse o segundo **no meio é que está a virtude**

Erro crasso exclamou o terceiro **no fundo do copo é que está o doce** Agarraram-se pelos cabelos e ainda não fizeram as pazes os danados Depois temos a história daquele Velho Provérbio que apeteceu-lhe uma pêra se pôs debaixo da árvore pensando **A pêra quando madura há-de cair** Mas a pêra só caiu quando já estava completamente podre e foi esborrachar-se em cima do tuitiço do Velho Provérbio que ofendido apresentou a sua demissão
Gianni Rodari, *Histórias ao Telefone* (texto adaptado), *Plural* 7

Este é um exercício de marcação de pontuação que inclui provérbios e também outras expressões idiomáticas não proverbiais. O negrito foi colocado por nós de forma a facilitar a identificação dos provérbios. O exercício apresenta algumas ajudas, uma vez que está assinalada a oposição entre maiúsculas e minúsculas, daí o aluno poder determinar mais facilmente onde deverá marcar as fronteiras de frase. Não está só em causa aqui o problema da pontuação, mas sobretudo a capacidade de distinção entre discurso direto e indireto e o conhecimento da forma como esta distinção é representada ortograficamente pela pontuação. Apenas o provérbio **Obra começada, meio acabada** apresenta rima interna. Todos os provérbios deste exercício apresentam uma estrutura bipartida e paralelismo entre os dois membros. A alteração da ordem dos constituintes em **no meio é que está a virtude e no fundo do copo é que está o doce** também poderá ser uma pista para a identificação deste tipo de estrutura. Neste excerto, a expressão **Quem age sozinho age por três** também se parece com um provérbio (cp. *Homem prevenido vale por dois*), apesar de não o reconhecermos como tal na nossa cultura nem de surgir nas coletâneas de provérbios consultadas ou na *internet*.

Este método permitiu-nos identificar ainda 34 provérbios que terão sofrido algum processo de reutilização e subversão estilística, distribuídos por 4 manuais escolares referentes aos 6º, 8º, 11º e 12º anos, com 17, 8, 4, e 5 provérbios, respetivamente. Abaixo, apresentamos alguns destes provérbios:

Casa onde não há sol todos ralham ou falam de futebol.

(cp. **Casa onde não há pão, todos ralham e ninguém tem razão.**)

Diz-me que cão tens, dir-te-ei quem és.

(cp. **Diz-me com quem andas, dir-te-ei quem és.**)

Quem semeia sol ilude tempestades.

(cp. **Quem semeia ventos colhe tempestades.**)

5. Resultados

Os dois métodos utilizados permitiram-nos encontrar 303 provérbios (272 provérbios

diferentes, uma vez que 31 provérbios foram captados quer pelos grafos quer pelos *scripts*). Na figura 3, apresentamos os resultados dos provérbios encontrados por ano escolar através dos diferentes métodos. Verifica-se que a maioria (71%) dos provérbios foi identificada pelo método dos termos e expressões introdutórias e apenas 18% pelas expressões regulares com comandos *grep*. 11% foram identificadas por ambos os métodos. O facto de o primeiro método ter apresentado melhores resultados pode estar ligado à natureza do *corpus*.

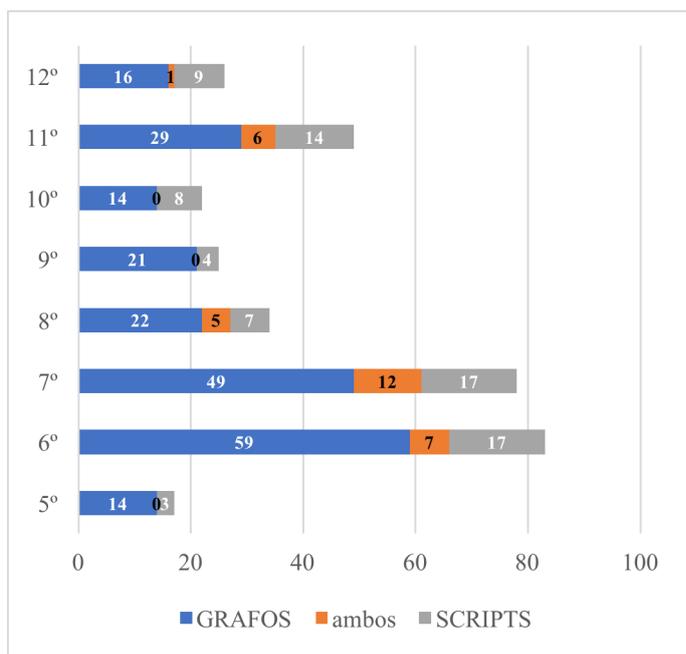


Figura 3, Provérbios encontrados por ano escolar

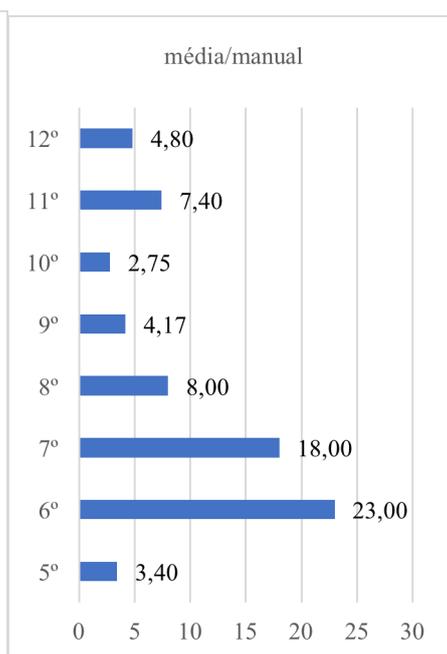


Figura 4, Média de provérbios por manual

Por outro lado, não obstante a distribuição do número de manuais por ano de escolaridade não ser homogénea verificou-se que existe uma alta correlação (coeficiente de Pearson: 0,97) entre o número de provérbios encontrados por ano escolar e a média de provérbios por manual e por ano letivo (Figura 4), ou seja, o número de manuais por ano letivo parece não ter influenciado o número de provérbios encontrados por ano escolar.

Ao procurarmos determinar quais os fatores que promovem a utilização dos provérbios nos manuais de Português, verificámos que estes tanto apareciam em textos como em diferentes tipos de exercícios. Na tabela seguinte, apresentamos a distribuição dos provérbios pelos manuais escolares e por estes dois tipos de contexto de ocorrência (exercícios/textos):

Tabela 1, Distribuição dos provérbios pelos manuais escolares

ano	Exercícios	Textos
5º	15	2
6º	60	16
7º	61	5
8º	27	2
9º	18	7
10º	15	7
11º	28	15
12º	6	19
total	230	73

Como se pode verificar, os provérbios ocorrem sobretudo nos exercícios (76%) e muito menos (24%) nos textos. Apenas nos manuais do 12º ano tal não sucede. A variação na proporção de provérbios entre os dois tipos de contexto de ocorrência também não parece significativa (Pearson: 0,39 ignorando os valores do 12º ano).

Analisando melhor os contextos de ocorrência dos provérbios, verificámos também que, nos manuais escolares, os provérbios aparecem associados a diferentes tipos de exercícios, nomeadamente os que pedem para:

- i. fazer corresponder duas partes de listas de provérbios;
- ii. classificar as orações do provérbio;
- iii. pontuar o provérbio;
- iv. justificar o emprego de pontuação no provérbio;
- v. sublinhar todos os (nomes / verbos, etc.) no provérbio;
- vi. reescrever o provérbio, subvertendo-o;
- vii. escolher o provérbio que poderia servir de título a um texto;
- viii. escolher de entre os provérbios apresentados aquele que sintetiza melhor o assunto do texto;
- ix. explicar o significado dos provérbios apresentados;
- x. demonstrar que o provérbio se pode aplicar ao texto;
- xi. redigir um texto partindo de um provérbio.

Relativamente ao número de ocorrências, não foi possível determinar quais os provérbios mais usuais uma vez que poucos se repetiram ao longo dos manuais. Apesar disso, verificámos que alguns provérbios se repetem no mesmo manual e, por vezes, o mesmo

provérbio pode estar associado a vários exercícios diferentes. Disso é exemplo o provérbio **Quem conta um conto acrescenta um ponto**, que encontramos várias vezes no manual do 7º ano *Sentidos*, em diferentes contextos, associado aos seguintes exercícios:

- i. O texto que vais ler começa com uma referência a um provérbio: “**Quem conta um conto acrescenta um ponto**”. Sabes o que é um provérbio? Explica o significado do provérbio transcrito.
[exercício: definição/explicação]
- ii. O provérbio “**quem conta um conto acrescenta um ponto**” possui um fundo de verdade indiscutível. [...] Na primeira parte do texto, a das reflexões, o narrador reconhece que o provérbio tem um fundo de verdade. Porquê? [exercício: explicação/comentário]
- iii. No provérbio “**Quem conta um conto acrescenta um ponto**”, podemos identificar duas orações. Delimita-as.
[exercício de gramática]

Este provérbio foi, além disso, o provérbio que mais se repetiu nos manuais escolares, tendo surgido 7 vezes.

6. Discussão

Concluimos que, apesar de só haver uma referência explícita a provérbios nos documentos ordenadores no 2º ciclo de estudos (5º e 6º anos de escolaridade), este tipo de expressões aparece disseminado ao longo de todo o percurso escolar, mas com frequências baixas (média de 7 provérbios por manual). Os provérbios surgem com maior frequência nos manuais do 6º e 7º anos.

Considerando que os provérbios mais usuais se encontram já representados nas coletâneas e dicionários de provérbios, muitos dos provérbios encontrados não são dos mais usuais, por exemplo, enquanto o provérbio *Quem conta um conto acrescenta um ponto* ocorre na *internet* cerca de 102 vezes⁷, o provérbio *Ao amigo molestar, nem a rir nem a brincar* [*Passa Palavra*, 6] só ocorre 4 vezes.

Observou-se também o uso de provérbios opacos, isto é, cuja interpretação não é literal ou que são dificilmente interpretáveis, ou ainda cujas condições pragmáticas de uso são pouco evidentes:

*Bem prega Maria em casa vazia.
Depois do burro morto, cevada ao rabo.
Duro com duro, não faz bom muro.
Em Abril, queima a velha o carro e o carril.*

Verificou-se que, para alguns exercícios, a escolha de provérbios como material linguístico a analisar é totalmente irrelevante; qualquer outra frase poderia servir o mesmo objetivo. O motivo da escolha pode dever-se ao facto de os provérbios serem frases curtas e de as listagens se encontrarem acessíveis.

Ora, a escolha do provérbio deve ser relevante quer para o conhecimento destas expressões e do seu uso, quer para o objetivo pedagógico dos exercícios/atividades. Os exercícios/atividades devem ser adequados quer ao nível de escolaridade, quer às realidades socioculturais dos estudantes (meio rural/urbano, religião, etc.). É necessário criar os materiais de base que permitam facilitar a seleção dos provérbios mais adequados aos objetivos pedagógicos a atingir tanto em língua materna como em língua não materna com base em informações de uso extraídas de *corpora* de textos reais, modernos, de grandes dimensões com uso de ferramentas de PLN para identificação automática de provérbios (e variantes) em textos (Reis, em preparação).

7. Trabalho futuro

Num trabalho futuro pretendemos alargar o âmbito deste trabalho e determinar quais os provérbios (e suas variantes) mais difundidos nos manuais de Português Língua Não Materna e refletir na forma como estes são utilizados em contexto didático.

Notas

1. O Ensino Básico compreende três ciclos: 1º Ciclo: 1º, 2º 3º e 4º Anos; 2º Ciclo: 5º e 6º Anos e 3º Ciclo: 7º, 8º e 9º Anos[.] O conjunto completo destes três graus de ensino constitui o nível de ensino básico. O Ensino Secundário é constituído pelo: 10º ano[.] 11º ano[.] 12º ano[.] in <http://cdp.portodigital.pt/educacao-e-formacao/ensino-basico-e-secundario> (última consulta em março de 2017).
2. Saramago, José. (1995) *Ensaio sobre a Cegueira*. 2ª ed., Lisboa: Editorial Caminho.
3. O autor utiliza, “a designação “desvio ligeiro” sempre que a UF apresenta uma pequena alteração como, por exemplo, a intromissão de uma palavra no conjunto que constitui a combinatória lexical [...] [considera] “desvio acentuado ou muito acentuado” quando, por exemplo, uma palavra da combinatória é substituída por outra” [...] (Pereira, 2015: 264-265).
4. <http://www.clul.ul.pt/pt/recursos/314-corpora-of-ple>.
5. Os *scripts* que permitiram construir as expressões regulares a partir das sequências de palavras-chave e retirar do corpus as frases em que esses padrões ocorrem foram elaborados pelo Prof. Nuno Mamede

(UL-IST & INESC-ID Lisboa), a quem os autores agradecem todo o apoio prestado.

6. en.wikipedia.org/wiki/Stop_words; Witten et al. (2011): 329, 387).
7. Usando a expressão completa *verbatim* e restringindo a busca ao domínio de topo “.pt” (consultado em março de 2017)

Agradecimentos

Os autores agradecem ao Professor Doutor Nuno Mamede (Universidade de Lisboa – IST e INESC-ID Lisboa) o seu apoio na elaboração dos *scripts* utilizados para este trabalho. Este trabalho foi parcialmente financiado pela FCT (ref. UID/CEC/50021/2013). A primeira autora agradece à FCHS UAlg o apoio para a participação no 10º *Colóquio Interdisciplinar sobre Provérbios* (ICP16), Tavira, Portugal, 6-13 de novembro de 2016.

Referências

- CLARK, A.; FOX, C. and LAPPIN, S. (eds.). (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. Wiley-Blackwell.
- BALULA, J.; MATOS, I; SILVA, A.; MELÃO, D.; AMANTE, S. CASTELO, A. (2013). *As repercussões das "Metas Curriculares" nos manuais escolares de Português do Ensino Básico*. Departamento de Ciências da Linguagem: Escola Superior de Educação de Viseu. disponível em <https://www.researchgate.net/publication/283425479>
- BAPTISTA, José A.; SANTIAGO, Ana; ALMEIDA, Dina; ANTUNES, Paula & GASPAR, Regina. (2011). *Programa de Português L2 para Alunos Surdos - Ensino Básico e Secundário*. Ministério da Educação.
- BUESCU, Helena C.; MORAIS, José; ROCHA, Maria Regina & MAGALHÃES, Violante F. (2015). *Programa e Metas Curriculares de Português no Ensino Básico*. Governo de Portugal - Ministério da Educação e da Ciência.
- CHACOTO, Lucília. (1994). *Estudo e Formalização das Propriedades Léxico-Sintáticas das Expressões Fixas Proverbiais*. Dissertação de Mestrado. Lisboa: Faculdade de Letras da Universidade de Lisboa (não publicada).
- CONSELHO DA EUROPA. (2001). *Quadro Europeu Comum de Referência para as Línguas: Aprendizagem, ensino, avaliação*. Porto: Edições ASA.
- DUARTE, Telma E. O. (2013). *Cinco unidades didáticas com provérbios: a produtividade dos provérbios na síntese e memorização de conhecimentos*. Relatório científico. Universidade do Porto.
- FERREIRA, Ana Filipa Antunes Alves. (2011). *O provérbio no ensino aprendizagem do PLE. Um pequeno passo para os alunos, um grande passo para uma competência comunicativa e intercultural*. Tese de Mestrado. Faculdade de Letras: Universidade do Porto.
- FERREIRA, Helena Maria & DE PAULA VIEIRA, Mauricéia Silva. (2013). O trabalho com o gênero provérbio em sala de aula. *Revista Línguas & Letras*, 14(26), 1-18
- LEIRIA, Isabel (coord.); MARTINS, Ana M.; CORDAS, Júlia; MOUTA, Margarida; HENRIQUES, Rosa. (2008). *Orientações Programáticas de Português Língua Não Materna (PLNM) Ensino Secundário*. Ministério da Educação.
- MARTINS, Pedro. (2010). Do provérbio em contexto didático: proposta de trabalho. *Paremia*, 19: 2010, pp. 93-102
- PAUMIER, S. (2016). *Unitex 3.1 - User Manual*. Université de Paris-Est/Marne-la-Vallée - Institut Gaspard Monge, Noisy-Champs.
- PELLEGRINI, T.; Ling, W.; Silva, A.; Correia, R.; Trancoso, I.; Baptista, J. and Mamede, N. (2012). Overview of Computer-assisted Language Learning for European Portuguese at L2F. In *Proceedings of the 4th International Conference on Computer Supported Education - CSEDU*: 538-543. Porto, Portugal.

- PEREIRA, António da Costa. (2015). *Competência Linguística em Português Europeu Língua Materna e Língua Não Materna: Aquisição, Ensino e Aprendizagem de Expressões Idiomáticas*. Tese de Doutoramento. Instituto de Letras e Ciências Humanas: Universidade do Minho.
- REIS, S. [em preparação] *Expressões proverbiais do português - Estudo da variação formal e das relações discursivas, identificação automática em textos e aplicações clínicas e didáticas*. Tese de Doutoramento. Faro: UALG.
- REIS, S. & BAPTISTA, J. (2016a). "Portuguese Proverbs: Types and Variants". in Gloria Corpas Pastor (ed). *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Geneva: Editions Tradulex, 208-217.
- REIS, S. & BAPTISTA, J. (2016b). LET'S PLAY WITH PROVERBS? - NLP tools and resources for iCALL applications around proverbs for PFL. in *Proceedings of the International Congress on Interdisciplinarity in Social and Human Sciences*, 5 a 6 de maio de 2016, University of Algarve, Faro, Portugal, 427-446.
- SCHNEIDER, G., North, B., Flügel, Ch. & Koch, L. (1999). *Europäisches Sprachenportfolio – Portfolio européen des langues – Portfolio europeo delle lingue – European Language Portfolio, Schweizer Version*. Berna: EDK. (Também disponível em: <http://www.unifr.ch/ids/portfolio>.)
- WITTEN, I.H.; FRANK, E. and HALL, M.A. (2011). *Data Mining Practical Machine Learning Tools and Techniques*. 3rd edition, USA: Morgan Kaufmann Publishers Inc.

Coletâneas de provérbios incluídas no corpus

- COSTA, J. (1999). *O Livro dos Provérbios Portugueses*. Lisboa. Editorial Presença.
- MACHADO, J.P. (1996). *O Grande Livro dos Provérbios*, 1^a edição, Lisboa: Editorial Notícias.
- MOREIRA, A. (1996). *Provérbios Portugueses*. Lisboa, Editorial Notícias.
- PARENTE, S. (2005). *O Livro dos Provérbios*. 1^a edição, Lisboa: Editora Âncora.

Curriculum Vitæ

Sónia Reis

Email: reis.soniamm@gmail.com

Sónia Reis é licenciada em Línguas e Literaturas Modernas - Estudos Portugueses e mestre em Ciências da Linguagem pela Universidade do Algarve, tendo também concluído a parte curricular do programa de doutoramento em Ciências da Linguagem da mesma universidade. A sua tese de doutoramento incide sobre a função discursiva das expressões proverbiais, isto é, a forma como estas se articulam com o discurso em que encontram, abordando também problemas de delimitação e identificação automáticas de provérbios em textos, usando ferramentas de processamento de linguagem natural, com o objetivo de desenvolver diversas aplicações didáticas e terapêuticas/de diagnóstico.

Jorge Baptista

Email: jbaptis@ualg.pt

Jorge Baptista é docente da Universidade do Algarve e investigador do INESC-ID Lisboa, no grupo do Laboratório de Língua Falada (L2F), tendo ampla experiência e numerosas publicações em Processamento Computacional de Linguagem Natural, especificamente em Português, incluindo o desenvolvimento de aplicações computacionais ao ensino da língua. Tem igualmente trabalho publicado relacionado com as ferramentas computacionais de diagnóstico e apoio terapêutico a diversas patologias da linguagem.