

VI SIMPÓSIO MUNDIAL DE ESTUDOS DA LÍNGUA PORTUGUESA

A UNIÃO NA DIVERSIDADE

Escola Superior de Educação de Santarém
Departamento de Línguas e Literaturas

SIMPÓSIO 77

O provérbio como estímulo num terapeuta virtual

Autores:

Sónia Reis

UALg, Faculdade de Ciências Humanas e Sociais, Departamento de Artes e Humanidades. reis.soniamm@gmail.com. Faro, Portugal.

Annamaria Pompili

INESC-ID Lisboa, Spoken Language Systems Lab (L2F). anna@l2f.inesc-id.pt. Lisboa, Portugal.

Alberto Abad

INESC-ID Lisboa, Spoken Language Systems Lab (L2F). alberto.abad@l2f.inesc-id.pt. Lisboa, Portugal.

Jorge Baptista

UALg, Faculdade de Ciências Humanas e Sociais, Departamento de Artes e Humanidades. Faro. Portugal & INESC-ID Lisboa, Spoken Language Systems Lab (L2F). jbaptis@ualg.pt. Lisboa, Portugal.

Resumo:

Os provérbios são elementos úteis no diagnóstico e terapia de certas patologias da linguagem, nomeadamente as que resultam de trauma, já que estão associados a estruturas da memória que são afetadas diferencialmente, mesmo quando a capacidade de falar é diminuída. Por esta razão, os provérbios têm vindo a ser utilizados como auxiliar de diagnóstico e de terapia, na medida em que constituem uma forma de exercício das estruturas cognitivas a partir da memória de longa duração. Apesar da sua ubíqua utilização por parte de terapeutas, a seleção dos provérbios para estímulo numa terapia ou num diagnóstico pode apresentar dificuldades, uma vez que não é fácil determinar se o não reconhecimento de um dado provérbio pelo paciente está associado a quadros patológicos ou ao mero desconhecimento do mesmo. Importa, por isso, que os provérbios e variantes a utilizar para estímulo sejam os de uso mais frequente e que possam ser reconhecidos pela generalidade dos falantes. Nesta comunicação, apresentamos a

metodologia seguida na construção de um módulo de exercícios de diferentes tipos envolvendo provérbios a integrar num Terapeuta Virtual para o tratamento da Afasia (VITHEA).

Palavras-chave: provérbios; afasia; terapeuta virtual; exercícios de diagnóstico e de terapia de afasia

1. INTRODUÇÃO

Os provérbios, devido à sua riqueza cultural e linguística, têm vindo a ser utilizados por diferentes especialistas e estão relacionados com as mais diversas áreas do conhecimento e da experiência humana. Estes são também elementos úteis no diagnóstico e terapia de certas patologias da linguagem, nomeadamente as que resultam de trauma, já que estão associados a estruturas da memória que são afetadas diferencialmente, mesmo quando a capacidade de falar é diminuída (Cazelato, 2003; Santos *et al.* 2009; Sé, 2011). Por esta razão, os provérbios têm vindo a ser utilizados como auxiliar de diagnóstico e terapia, na medida em que constituem uma forma de exercício das estruturas cognitivas a partir da memória de longa duração. Estes exercícios consistem, basicamente, em pedir aos pacientes que completem o provérbio, digam qual o seu significado ou expliquem o contexto em que este pode ser utilizado.

Apesar da sua utilização por parte de terapeutas, a seleção dos provérbios pode apresentar dificuldades, uma vez que não é fácil determinar se o não reconhecimento de um dado provérbio pelo paciente está associado a quadros patológicos (p. ex. demência) ou ao mero desconhecimento do mesmo. Importa por isso que os provérbios e variantes a utilizar para estímulo sejam os de uso mais frequente e que possam ser reconhecidos pela generalidade dos falantes.

Tendo estes aspetos em consideração, um dos objetivos deste trabalho é desenvolver um módulo de exercícios de diferentes tipos, construídos com base em provérbios portugueses usuais, com vista a integrá-los num terapeuta virtual, o sistema VITHEA (Abad *et al.* 2013).

Este artigo está organizado do seguinte modo: em seguida, apresentamos brevemente alguns trabalhos em terapia da fala que recorrem a provérbios sobretudo para diagnóstico de diversas patologias da linguagem, apresentando de seguida o terapeuta virtual VITHEA em que iremos integrar os resultados na nossa investigação. Apresenta-se depois a metodologia empregue na determinação dos

provérbios usuais, a tipologia de exercícios a construir e os critérios de seleção de distratores. Finalmente, apresentamos o inquérito realizado para validar esta metodologia e discutimos os resultados obtidos. Concluímos com algumas perspetivas de trabalho futuro.

Vários provérbios referem os provérbios quer como elemento de diagnóstico, quer como de terapia (Pies, 1994; Gorham, 1995; Kent, 2004; Morato, 2010). Ainda que alguns destes trabalhos se refiram à variante brasileira do português, o interesse pelo uso de provérbios neste contexto pode ser generalizado, ainda que com cautela, na medida em que existe um repertório comum às variantes europeia e brasileira. Não encontramos, porém, trabalhos que justificassem a seleção de provérbios apresentada nos instrumentos de diagnóstico ou terapia desenvolvidos.

No diagnóstico de certas patologias da linguagem, que podem resultar de diferentes fatores, Cazelato (2003) teve como objetivo o estudo do processo de significação que ocorre na interpretação e enunciação linguístico-discursiva de provérbios por sujeitos afásicos, analisando o percurso enunciativo realizado pelos sujeitos aquando do reconhecimento ou tentativa de reconhecimento de determinado provérbio e o do seu "correspondente" semântico-pragmático, que denomina por "provérbio equivalente". Dito de outro modo, foi observada a forma como os sujeitos afásicos relacionam o provérbio com o seu possível significado e o associam a outro provérbio com um significado semelhante. Para tal, foi utilizado o *Protocolo de Estudo de Provérbios Equivalentes*, que é composto por 57 provérbios.

Não encontrámos explicitada a forma como a autora selecionou, a partir do património paremiológico da língua, os provérbios utilizados nos protocolos que construiu. É referido que os provérbios são apropriados à configuração sócio-cultural e à realidade dos sujeitos em estudo; que foi considerada a distinção entre provérbios que têm presente uma metáfora e provérbios que têm um sentido mais literal e que foram selecionados os provérbios-alvos e os provérbios equivalentes, considerando os seus significados.

Santos (2009) ao constatar a escassez de instrumentos para a avaliação e triagem da doença de Alzheimer (DA), elaborou e validou o *Teste de Rastreo da Doença de Alzheimer com Provérbios (TRDAP)*, a partir de um Jogo de Memória de provérbios e tarefas a serem realizadas que visam avaliar a Memória de Curto

Prazo (MCP), as Funções Executivas e Linguagem (FE e L) e a Memória Episódica (ME) (Stemmer & Whitaker, 2008) dos indivíduos a partir dos sessenta anos. Este teste demonstrou ser um instrumento facilitador do rastreio para os profissionais e um auxílio na reabilitação cognitiva dos adultos ou idosos com DA. A autora não indica como foram selecionados os provérbios incluídos no teste que construiu.

Também Sé (2011) desenvolveu um estudo sobre a interpretação e manipulação enunciativa dos sentidos veiculados nos provérbios por sujeitos com provável doença de Alzheimer leve. Também nesse estudo, a escolha dos provérbios não foi justificada, uma vez que teve por base o *Protocolo* proposto por Cazelato (2003).

É de grande importância que os provérbios que fazem parte deste tipo de estudo sejam expressões usuais. Só assim é que se pode determinar se o não reconhecimento de um provérbio pelo paciente se deve ou não a fatores que estão associados a certas patologias da linguagem. No nosso trabalho, pretende-se melhorar este aspeto, contribuindo com informação de natureza quantitativa para auxílio à seleção de provérbios para fins de diagnóstico e/ou terapêutica, nomeadamente por meio da definição dos índices de frequência dos provérbios o que possibilitará a construção de elementos de diagnóstico mais fiáveis.

O VITHEA (Virtual Therapist for Aphasia Treatment)¹ é um programa de reabilitação da linguagem, vocacionado para as dificuldades anómicas de expressão oral de doentes afásicos. A afasia é uma perturbação adquirida da linguagem após uma lesão cerebral, sendo que a sua etiologia mais comum é o acidente vascular cerebral. A afasia caracteriza-se por dificuldades de expressão oral, de compreensão auditiva, de leitura e de escrita, embora a manifestação mais persistente e comum seja a dificuldade de evocar o nome de um determinado objecto. A terapia da fala tem demonstrado ser eficaz no tratamento da afasia (Basso, 1992; Mazzoni, 1995) sendo confirmada em estudos de imagem funcional (Small *et al.*, 1998; Musso *et al.*, 1999; Léger *et al.*, 2002; Peck *et al.*, 2004), salientando-se a importância da intensidade do tratamento (Bhogal *et al.*, 2003).

As atuais tecnologias da língua podem ter um papel muito relevante no desenvolvimento de métodos automáticos de terapia que permitam um acesso frequente e intensivo às sessões de reabilitação. O VITHEA é uma plataforma on-

¹ <https://vithea.l2f.inesc-id.pt/>

line projetada para agir como um "terapeuta virtual", apresentando exercícios de nomeação aos pacientes através de um agente animado, que desempenha o papel de terapeuta virtual. Concretamente, o sistema integra a tecnologia de reconhecimento automático de fala para validar as respostas dos pacientes. A solução adotada é baseada numa abordagem que, dado um fluxo contínuo de áudio, procura apenas as palavras-chaves correspondentes às possíveis respostas esperadas. Desta forma o sistema torna-se bastante robusto, o que é fundamental para o paciente desenvolver estratégias de autocorreção. Após validação da resposta, o sistema fornece um feedback tanto escrito como oral, através de mensagens de voz produzidas pelo terapeuta virtual usando um sintetizador de fala.

2. METODOLOGIA E RESULTADOS

O primeiro desafio deste estudo consistiu na definição de um corpo de provérbios suficientemente usuais para que pudesse ser usado, com fiabilidade, na elaboração de exercícios. Para a constituição do *corpus* de trabalho, partiu-se de uma base dados com mais de 114 mil provérbios reunidos a partir de quatro coletâneas, que foram digitalizadas e convertidas em formato de texto para processamento automático (Reis & Baptista, 2016).

2.1. Determinação dos provérbios usuais

O grau de generalidade dos provérbios foi avaliado através de um inquérito por questionário² (Reis & Baptista, 2017b), difundido via *web* com recurso à ferramenta *Google Forms*³.

Embora dispéssemos de uma listagem muito extensa de provérbios, com mais de 114 mil entradas, não seria possível aferir para cada um deles e por meio de um questionário se era ou não de uso geral, pelo que se utilizou apenas uma amostra aleatória, constituída de forma semiautomática a partir desta listagem, incluindo não só provérbios usuais como provérbios pouco usuais, e assim validar a classificação feita nessa amostra recorrendo a um número elevado de respondentes. Dado o tempo que a aplicação deste tipo de instrumentos implica

² <https://goo.gl/forms/PH8uMlnhsFMlz2SQ2>

³ <https://www.google.com/forms>.

para cada respondente, decidiu-se que o número de provérbios a apresentar seria apenas de 100, metade usuais e a outra metade pouco usuais; o respondente deveria selecionar uma das três hipóteses: *conheço e uso*; *conheço, mas não uso*; *não conheço*.

A seleção manual dos provérbios foi feita independentemente por dois investigadores, a partir da listagem com mais de 114 mil provérbio. Cada um indicou os provérbios ou variantes de provérbios que considerou mais usuais. Ao todo, os 2 anotadores assinalaram como usuais 739 e 379 provérbios (ou suas variantes) por cada um dos anotadores. De seguida, as duas sublistagens foram cotejadas. Os provérbios assinalados por ambos os anotadores receberam a classificação '2', os que apenas foram marcados por um dos anotadores foram classificados com '1'. Todos os restantes foram marcados com '0'. Ao todo, foram assinalados 276 provérbios (e variantes) de tipo '2' e 566 de tipo '1'. A intuição inicial era que os provérbios de nível 2 eram *bastante usuais*, os de nível 1 eram menos *usuais*, mas ainda bem conhecidos da generalidade dos falantes e os de nível 0 eram *não usuais*. Depois de ordenados aleatoriamente, selecionaram-se para o questionário os primeiros 25 provérbios do nível '2', outros tantos do nível '1' e 50 do nível '0'. Teve-se o cuidado de colapsar as diferentes variantes repetidas do mesmo provérbio numa só, escolhendo, após breve troca de impressão entre os anotadores, a variante que pareceu mais usual. Os provérbios foram apresentados de forma aleatória aos inquiridos.

O inquérito foi disponibilizado no dia 6 de fevereiro de 2017 e esteve online até ao dia 12 do mesmo mês, tendo recebido 735 respostas. Apesar do número de respostas ser expressivo, não é possível considerar a amostra representativa da população portuguesa no seu todo. Na tabela 1, apresentamos a matriz de confusão das respostas fornecidas pelos informantes quando comparadas com os níveis de disponibilidade lexical propostos pelos 2 anotadores. A tabela interpreta-se da seguinte maneira: na primeira célula indica-se a percentagem do nº de respostas em que os informantes assinalaram os provérbios com "não conheço", provérbios que estavam classificados como de nível 0. Por outras palavras tantos os informantes como os anotadores concordam de que se tratam de provérbios pouco usuais. Com as devidas alterações, o resto da tabela interpreta-se da mesma maneira.

Tabela 1. Matriz de confusão entre respostas dos inquiridos e níveis de disponibilidade lexical, em percentagem (N=735 respondentes x 100 respostas)

	<i>n0</i>	<i>n1</i>	<i>n2</i>
<i>r0</i>	0,41031	0,07400	0,01569
<i>r1</i>	0,06137	0,08086	0,10777
<i>r2</i>	0,02792	0,05707	0,16501

Este questionário permitiu-nos confirmar, em geral, os níveis de disponibilidade lexical atribuídos aos provérbios anotados como pertencendo aos níveis 2 e 0. Verificámos que os respondentes têm mais tendência a acertar se os fatores forem *conhece* e *não conhece* os níveis de disponibilidade lexical dos níveis 2 e 0. Desta forma, considerando apenas a oposição dos provérbios de nível 0 que foram reconhecidos pelos respondentes; e os provérbios considerados conhecidos independentemente de usados ou não pelos informantes (de nível 0 e de nível 2) 82% das respostas confirmam a nossa classificação inicial. Os casos dos provérbios de nível 1 devem ser considerados com cuidado, pois os resultados não mostram uma distinção clara entre esse nível e os outros dois.

A lista de provérbios usuais já serviu de referência para diversas observações sobre o uso de provérbios em textos (Reis & Baptista, 2016b; 2017a; 2017b), bem como para a elaboração dos exercícios, que apresentaremos de seguida. Um módulo de exercícios já foi avaliado através de um inquérito por questionário apresentado no ponto 2.4 deste trabalho.

2.2. Jogos - tipologia

Os questionários com perguntas de escolha múltipla *cloze questions* têm sido uma estratégia popular de ensino-aprendizagem, descrita desde, pelo menos, Taylor (1953). Este tipo de questionário consiste na remoção de uma ou mais palavras de um enunciado (estímulo), que deverão ser preenchidas com base no conhecimento prévio ou experiência pessoal do respondente, a partir de um conjunto finito de hipóteses de respostas. Nestas, a par da resposta certa, figuram um conjunto de

distratores, cuja seleção segue, muitas vezes, princípios decorrentes dos objetivos de aplicação do questionário.

Estes questionários têm sido utilizados para mobilizar diferentes tipos de conhecimento, inclusivamente no ensino de línguas. Vários trabalhos foram desenvolvidos em torno dos métodos de construção de perguntas de escolha múltipla, de forma a permitir uma eficiente eliciação do conhecimento controlando eventuais enviesamentos nas respostas decorrentes da forma como as questões são apresentadas. Alguns desses estudos visam ainda a sua construção automática (Gangur, 2011), incluindo a construção automática dos distratores (Correia *et al.* 2010).

Em Reis & Baptista (2016b) fez-se um levantamento dos tipos de exercícios já existentes, construídos em torno de provérbios, e focados essencialmente no ensino de línguas. Verificou-se que uma grande parte dos exercícios/jogos já disponíveis consistia basicamente na tarefa de completar um provérbio, do qual se apresentava apenas um dos hemistíquios ou é removida uma ou mais palavras. Estes exercícios permitem treinar diversas competências linguísticas, além de promoverem o melhor conhecimento do provérbio.

Numa primeira fase, tendo esta tipologia de exercícios em mente, foi desenvolvido manualmente um conjunto de exercícios, com base nos provérbios mais usuais, definidos pelo processo descrito em 2.1. Entre outros objetivos, o estabelecimento deste conjunto de provérbios usuais visou, por um lado, a sua integração na plataforma virtual VITHEA, um terapeuta de fala virtual para utilizadores com diferentes tipos afasia; por outro lado, o desenvolvimento de jogos didáticos a aplicar em ensino de língua, especialmente em PLE. Os critérios de seleção de distratores mereceu especial atenção, com o intuito de, posteriormente, estes exercícios poderem vir a ser produzidos (semi)automaticamente, com base nos recursos linguísticos existentes para o processamento computacional do português.

Um exercício com o mesmo provérbio poderá ser de diferentes tipologias. A escolha do tipo de exercício a realizar dependerá do perfil dos destinatários, nomeadamente a sua patologia ou o objetivo de aprendizagem a atingir em L2. São exemplos de exercícios possíveis: seleção de uma imagem adequada ao provérbio; comparar a resposta oral do paciente com a resposta certa, que corresponde à

palavra-alvo omitida no provérbio (como já se faz no VITHEA, com cerca de 70 provérbios); escolher de entre um conjunto de palavras apresentadas a que corresponde à palavra-alvo no provérbio (como pretendemos implementar). Apesar de a plataforma VITHEA já apresentar um número interessante de exercícios com provérbios, e que podemos verificar tratar-se, na sua grande maioria de expressões usuais, deixa, no entanto, de fora numerosas outras unidades paremiológicas, cujo uso é de grande generalidade, e que pretendemos com este trabalho complementar.

Neste trabalho, porém, desenvolvemos um módulo de exercícios que consistem em selecionar, de entre um conjunto de respostas apresentadas, a que melhor completa o provérbio que serve de estímulo e que é apresentado através de um enunciado escrito. Note-se que este tipo de exercício apenas se destina a utilizadores sem perturbações da linguagem que interfiram no processo de leitura, ou seja, pessoas que têm a capacidade de ler e interpretar adequadamente um enunciado escrito. A resposta consiste apenas na seleção da opção considerada correta, não requerendo à competência da escrita nem à resposta oral (como faz atualmente o VITHEA)⁴.

Para cada estímulo (provérbio) foi desenvolvido um conjunto de 3 distratores (opções de resposta) linguisticamente motivados e que assentou nos critérios apresentados a seguir, em 2.3. Foram selecionados 100 provérbios de nível 2 (muito usuais) e construídos os estímulos correspondentes, distribuídos de igual forma, com base em dois tipos de exercício: (i) com a remoção de uma palavra-alvo e (ii) com a remoção de um dos membros do provérbio (o primeiro ou o segundo).

2.3. Critérios para a seleção dos *distratores*

A seleção dos *distratores* obedeceu a diversos critérios, que assentam essencialmente em reflexões linguísticas, e tem como ponto de partida as propriedades linguísticas características dos provérbios. Desta forma, tendo em

⁴ Neste sentido, é possível que o módulo aqui apresentado venha a integrar um sistema semelhante ao VITHEA-Kids (Mendonça *et al.* 2015), um desenvolvimento do projeto VITHEA original, com cuja tipologia de exercícios tem maior afinidades.

consideração a estrutura formal, bem como as propriedades sintáticas, semânticas, lexicais, fonéticas dos provérbios, foram desenvolvidos *distratores* que apresentam diferentes relações com as palavras-chave dos mesmos. Essas relações são essencialmente de natureza semântica (hiperonima/hiponímia, holonímia/meronímia, sinonímia/antonímia), paralelismos sintáticos e relações fonéticas.

Dezasseis dos *distratores* deste inquérito não seguem este padrão, não estabelecendo qualquer relação entre as palavras-alvo dos provérbios. Estes designamos por *distratores non sense*. Por exemplo, para os seguintes provérbios, escolheram-se os seguintes *distratores*:

(i) *Palavra puxa* [*palavra*]; *distratores*: (1) *argumento*, (2) *frase*, (3) *carroça*;

(ii) Em *Todos os caminhos* [*vão dar a Roma*]: *distratores*: (1) *levam a Roma*, (2) *vão dar a Oklahoma*, (3) *vão dar ao rio*.

No primeiro caso, a opção *carroça* é o *distrator non sense*, embora explore a semelhança com o provérbio *Peixe não puxa carroça*. No segundo caso, explora-se a semelhança com o provérbio *Todos os rios vão dar ao mar*. Assim, mesmo este tipos de *distratores* obedeceu a critérios de seleção, procurando sempre motivar a sua escolha para o exercício.

Para 12 provérbios indicamos entre as opções de resposta não apenas uma mas duas soluções certas. A escolha deste critério deveu-se a diversos fatores: por um lado, seria interessante verificar qual a escolha do respondente quando este conhecesse mais do que uma opção correta, uma vez que este questionário só permite a seleção de uma das opções (pensou-se que a escolha recairia sobre a variante mais comum do provérbio); por outro lado, nos casos em que os respondentes não conhecessem os provérbios, verificar-se-ia quais as diferenças entre a taxa de acerto dos provérbio com uma ou mais do que uma respostas certas.

3. RESULTADOS E DISCUSSÃO

Para avaliar a qualidade dos jogos construídos foi aplicado um inquérito por questionário usando a ferramenta *Google Forms*. Foram selecionados 100 provérbios de *nível 2* e construídos os respetivos exercícios obedecendo a dois

tipos de questão, consoante se trata de completar uma palavra ou todo um membro do provérbio. O questionário está dividido em três secções: na *secção 1* são requeridos os dados pessoais dos respondentes (sexo, idade, nacionalidade, escolaridade e distrito habitual de residência; na *secção 2* são apresentados os provérbios a completar; na *secção 3* são apresentadas 5 questões ao respondente sobre o próprio questionário e é disponibilizado um espaço para que este possa deixar a sua opinião e o seu contacto de e-mail, caso pretenda receber mais informações sobre o projeto. Tanto as perguntas como o conjunto de respostas a cada pergunta foram apresentadas aleatoriamente.

Por diversas razões práticas, recorreu-se a uma amostra de conveniência para a seleção dos respondentes, recorrendo a contactos informais e distribuindo o inquérito via e-mail e através das redes sociais, como explicaremos de seguida. A distribuição deu-se através de 4 grupos de contactos pessoais no *Facebook* (cerca de 3.488, muito diversificados) e através de uma lista de cerca de 150 contactos pessoais do *Gmail* (com uma média etária dos 30-50). A taxa de resposta estimada com base no total dos grupos de contactos pessoais é de cerca de 4%. O inquérito foi disponibilizado no dia 12 de outubro de 2017 e esteve *on-line* até ao dia 17 do mesmo mês, tendo recebido 161 respostas. De seguida, caracterizamos sucintamente a amostra. Os sujeitos da amostra são maioritariamente do sexo feminino (74%), atestando uma grande assimetria de género dos respondentes. Destes, 84 (52%) têm idades compreendidas entre os 31 e os 50 anos, 39 (24%) têm mais de 50 anos, 36 (23%) têm idades entre os 18 e os 30 anos e somente 2 respondentes têm menos de 18 anos. Os inquiridos são de 4 nacionalidades diferentes, sendo a maioria (150) de nacionalidade portuguesa, correspondendo a 93% da amostra. De acordo com os dados recolhidos, 138 indivíduos possuem o ensino superior (86). Quanto à residência, os respondentes indicam 9 distritos de Portugal Continental, maioritariamente em Faro (92, 58%), Lisboa (35, 22%) e Porto (10, 6%); e apenas 12 dos inquiridos referiram que residiam fora de Portugal (tendo as suas respostas sido removidas da amostra).

Note-se que estes dados não, pois, ser considerados representativos da população portuguesa no seu todo e que se deveram, sobretudo, à forma como foi distribuído o questionário. As assimetrias mais acentuadas dizem sobretudo respeito ao local de residência dos inquiridos e à sua escolaridade. De acordo com

os resultados dos Censos 2011⁵, mais de 80% da população residente em Portugal se concentrava nas regiões do Norte (34,95), Lisboa (26,7%) e Centro (22,0%). Faro (Algarve), apesar de ser o distrito de onde se obteve o maior número de respostas no nosso inquérito, apenas representava 4,3% da população. No que respeita ao nível de escolaridade, e ainda segundo os mesmos censos, apenas 15% da população com 23 ou mais anos tem ensino superior.

De forma a conseguirmos obter uma amostra o mais homogénea possível, optámos, nesta fase do trabalho, por analisar apenas os dados referentes aos informantes de nacionalidade portuguesa e que residem em Portugal. Isto deveu-se a poder haver influência no grau de acerto quer pelo facto de as pessoas residirem fora do país (não se perguntou há quanto tempo), quer pelo facto de não terem nacionalidade portuguesa, o que poderá levar a que o seu conhecimento de português e da cultura portuguesa não seja equivalente ao de um falante nativo. Apesar de a maioria dos respondentes estrangeiros terem nacionalidade brasileira, decidimos, neste momento, ignorar estas respostas devido a potenciais diferenças entre as variantes de cada país. Assim, os resultados apresentados doravante dirão respeito a 146 respondentes.

Em média houve uma taxa de acerto de 93%, considerando apenas uma resposta certa por questão. Levando em conta os casos em que há duas respostas certas, a taxa de acerto passa a 94%.

A distribuição da taxa de acerto atingida por cada uma das 146 pessoas que retivemos na amostra, com *bins* de 5%, revelou que 69 pessoas (47%) tiveram entre 95 e 100% das respostas certas e 60 pessoas (41%) atingiram entre 90 e 94% de respostas certas; já no bin de taxas de acerto entre 85 e 89%, só encontramos 11 pessoas (8%). Os valores mais baixos de taxa de acerto foram de 53% (1 pessoa) e 78% (2 pessoas) e nenhuma pessoa acertou 100% das respostas.

Na construção do inquérito surgiram alguns problemas que poderão ter influenciado estes resultados. Houve um lapso numa das questões em que não se indicou a palavra-alvo entre as opções de resposta, facto que foi notado por vários respondentes. Assim, descartámos esta questão.

Verificou-se que para um provérbio uma opção incorreta (distrator) era selecionado num número elevado de respostas. Trata-se do provérbio [Q-090]

⁵ www.censos2011.ine.pt

Deus dá o frio [conforme a roupa] em que se omitira o segundo membro. As opções de resposta apresentadas e respetivo número de respostas foram as seguintes: (a) *e o Diabo dá o brio* (49%); (b) *conforme a roupa* (resposta certa: 31%); (c) *conforme o arrepio* (15%); e (d) *conforme a vestimenta* (5%). A maioria dos respondentes escolheram a resposta (a), que assenta no paralelismo sintático e rima. Neste caso, o nível de disponibilidade lexical deverá ser revisto, pois o provérbio parece não ser conhecido pela maioria dos respondentes.

Houve também outros quatro casos que se destacaram pelo facto de a taxa de acerto apresentar valores mais baixos, entre os 52 e os 64%. Trata-se dos provérbios: [Q-037] *A experiência é a mãe da ciência*, em que a taxa de acerto é de 64%; [Q-079] *Para quem é, bacalhau basta*, com 62%; [Q-087] *As moscas apanham-se com mel e não com fel*, também com 62% de taxa de acerto, e o provérbio [Q-018] *A má ação fica com quem a pratica*, com apenas 52% dos inquiridos a optar pela resposta certa. No caso do provérbio Q-087, pensa-se que houve uma interferência com a variante negativa *Não é com vinagre que se apanham moscas*, uma vez que 23% dos respondentes escolheram a opção *com vinagre*. O mesmo se verifica com o provérbio Q-018, onde 36% das respostas selecionadas correspondem à opção *boa*. Tal poderá resultar de uma interferência entre esta variante do provérbio e uma outra variante, atestada: *A boa ou má ação fica com quem a pratica*.

No final do questionário pedia-se aos inquiridos que respondessem a 5 questões a fim de caracterizar a sua perceção sobre o conhecimento prévio que tinham dos provérbios que lhe foram apresentados. A maioria dos inquiridos (86%) conhecia, na maior parte das vezes, os provérbios e 11% responderam que os conheciam todos. Quanto às opções indicadas, a maioria (54%) confirmou que a resposta certa se encontrava *sempre* entre as soluções, mas 44% que apenas se encontrava *na maioria das vezes*. À questão sobre se tinha conseguido completar o provérbio facilmente, com base nas respostas apresentadas, *mesmo quando não conhecia o provérbio*, 102 (70%) inquiridos responderam *a maioria das vezes* e 30 (21%) *sempre*. (Falaremos dos casos mais complexos com duas respostas certas já adiante). Finalmente, a maioria dos inquiridos (52%) indica que *usa a maioria dos provérbios* apresentados, embora 47% indique que *apenas usa alguns*.

As respostas quanto ao conhecimento prévio do provérbio e seu uso confirmam, em primeiro lugar, a adequada classificação dos provérbios apresentados como *muito usuais*. Note-se que a maioria dos provérbios incluídos neste inquérito não tinha feito parte do questionário que pretendia avaliar a disponibilidade lexical de provérbios e que foi apresentado em Reis & Baptista, 2017b, resultando a sua classificação como muito usuais apenas da anotação manual por dois linguistas. Estes novos resultados reforçam, pois, as conclusões a que chegámos naquele trabalho.

As respostas quanto à presença da solução certa entre as opções de resposta, bem como o uso dessas opções para encontrar a resposta certa, mesmo quando não conhecia o provérbio, valida a metodologia seguida, na seleção dos distratores, tendo em vista uma eficiente construção dos exercícios.

Tal como dissemos atrás, para 12 provérbios apresentaram-se, não apenas uma, mas duas respostas certas, uma que corresponde à variante mais usual e outra que consideramos menos usual, mas ainda assim bem conhecida da generalidade dos falantes. Trata-se de confirmar esta classificação das variantes, bem como verificar se os falantes têm consciência do fenómeno da variação, nomeadamente através de uma pergunta feita no final do questionário. Os resultados são apresentados na tabela seguinte, por ordem crescente de respostas em que foi escolhida a variante considerada menos usual (RC2).

Como se pode ver, na maioria dos provérbios (7), a resposta recaiu sobre a opção que consideramos ser a variante mais usual. No caso do provérbio Q-026 A [noite] *é boa conselheira*, verificou-se que a maioria das respostas (52%) também selecionou a resposta certa (noite), mas as restantes (47%) recaíram sobre a variante (almofada). No provérbio Q-008, O [tempo] *voa*, 18% dos inquiridos selecionaram a variante *dinheiro*. Já no caso do provérbio Q-044, [Palavras], *leva-as o vento*, 12% das respondentes escolheram a variante *cantigas*. Note-se que o distrator *canções* nunca foi selecionado. Finalmente, no provérbio Q-096, [Grão a grão] *enche a galinha o papo*, a variante *bago a bago* só foi escolhida por 8% dos respondentes.

Tabela 2: Taxa de acerto nas questões com duas respostas certas

Questão	R1	R2	R3	R4	RC1	RC2	Prec
[Q-005] A ___ da vizinha é sempre melhor que a minha.	0,00	0,01	0,97	0,01	0,97	0,01	0,99
[Q-092] Quanto mais alto se sobe, ___.	0,01	0,03	0,00	0,96	0,96	0,03	0,99
[Q-048] O silêncio é a alma do ___.	0,01	0,01	0,04	0,94	0,94	0,04	0,98
[Q-053] Todos os caminhos ___.	0,00	0,00	0,05	0,95	0,95	0,05	1,00
[Q-033] Burro velho não aprende ___.	0,01	0,06	0,00	0,93	0,93	0,06	0,99
[Q-096] ___ enche a galinha o papo.	0,08	0,92	0,00	0,00	0,92	0,08	1,00
[Q-044] ___, leva-as o vento.	0,12	0,88	0,01	0,00	0,88	0,12	0,99
[Q-022] Abril, ___ mil.	0,00	1,00	0,00	0,00	1,00	0,00	1,00
[Q-052] Diz-me com quem andas, ___.	0,99	0,00	0,01	0,00	0,99	0,00	0,99
[Q-008] O ___ voa.	0,18	0,00	0,82	0,00	0,82	0,18	1,00
[Q-099] Quem espera, ___.	0,76	0,24	0,00	0,00	0,76	0,24	1,00
[Q-026] A ___ é boa conselheira.	0,00	0,01	0,52	0,47	0,52	0,47	0,99

Destes resultados podemos inferir que as variantes recenseadas dos provérbios apresentados, ainda que usuais, parecem ter uma disponibilidade lexical menor do que a variante que considerámos como mais usual. A exceção parece ser a das variantes *noite/almofada* do provérbio Q-026, cuja repartição é praticamente idêntica.

Notamos ainda o caso do provérbio Q-099, *Quem espera [sempre alcança]*, em cujo conjunto de opções de resposta incluímos *desespera*, que foi selecionado por 24% dos inquiridos. Na realidade, trata-se de um outro provérbio que tem o mesmo início, mas cujo significado e condições pragmáticas de uso é perfeitamente distinto. Ainda assim, um dos provérbios parece ter uma disponibilidade lexical maior do que o outro.

A existência de questões com mais de uma resposta certa não foi percebida por 79 dos inquiridos (54%) ou apenas foi percebida como existindo *poucas vezes* por 62 pessoas (42,5%); 4 dos inquiridos indicaram haver mais do que uma resposta certa *na maioria das vezes* e apenas 1 indicou que tal sucedia *sempre*.

Tal revela que a maioria das pessoas apenas conhece uma variante do provérbio, embora uma percentagem muito significativa tenha consciência do fenómeno da variação.

Como já se referiu anteriormente, uma parte do questionário foi disponibilizada para que os respondentes pudessem dar a sua opinião sobre o próprio questionário. Foram obtidas 22 opiniões das quais vamos comentar

algumas: Oito dos respondentes observaram que na questão [Q-006], referente ao provérbio *São mais as [vozes] que as nozes*, não fora apresentada a opção correta (*vozes*). Como dissemos acima, os dados relativos a esta questão foram removidos.

Quatro respondentes de nacionalidade brasileira referiram que tiveram dificuldades em responder ao questionário, pois não conheciam todos os provérbios ou as variantes apresentadas. Um destes salientou ainda que gostaria que o questionário não tivesse todas as questões de carácter obrigatório, uma vez que essa circunstância o levou a “responder às cegas”. Este tipo de comentário fundamentou a decisão de remover da amostra as respostas de inquiridos de nacionalidade estrangeira, a que aludimos atrás.

Um respondente referiu que tinha encontrado duas respostas certas num dos provérbios (Q-099) e que gostaria de poder assinalar ambas as opções, mas o sistema não permitia. Tal corresponde a uma situação que pretendíamos analisar especificamente e que tratámos atrás.

Um outro respondente mencionou que, no final do questionário, deveriam ser dadas as opções corretas das respostas para que ficasse a conhecer se respondeu ou não corretamente. Este não era, efetivamente, um dos objetivos deste questionário. Findo o processo de análise, todos os inquiridos que facultaram o seu contacto receberão a respetivas soluções.

Embora um dos inquiridos tenha comentado que as respostas certas estavam maioritariamente na 1ª ou 2ª opção, na realidade a distribuição da posição da resposta certa entre as outras opções de resposta, para cada pergunta, foi a seguinte: (R1, como primeira opção) 20 questões, (R2) 30 questões, (R3) 32 questões e (R4) 17 questões (considerando apenas uma resposta certa por questão). Apesar do número ligeiramente maior de perguntas com respostas certas na posição R2 e R3, a distribuição é bastante equilibrada. Uma vez que as questões foram também apresentadas aleatoriamente, julgamos que tal fator não terá influenciado a realização do inquérito.

4. CONCLUSÕES E TRABALHO FUTURO

Neste estudo, apresentamos a metodologia seguida na elaboração de um módulo de exercícios construídos com base em provérbios usuais, tendo em vista a sua

integração num terapeuta virtual, mas que poderá igualmente ser adequado para uso em contextos didáticos. A definição do conjunto de provérbios usuais, a partir de um corpus de mais de 114 mil provérbios, foi apresentada e validada experimentalmente quer com base em dados resultantes de inquérito quer com recurso a pesquisa na internet e em corpora (Reis e Baptista 2017b). Com base nesta listagem, descrevemos os critérios de seleção de distratores para a elaboração de questionário, tendo-o validado com base numa mostra de 99 respostas, resultante de inquérito a 161 respondentes. Os resultados confirmam que os critérios adotados para a seleção dos distratores dos provérbios são adequados, verificando-se que a maioria dos respondentes (88%) obtém uma elevada taxa de acerto (acima de 90%). Alguns dos critérios adotados para a construção do questionário, nomeadamente na seleção dos distratores, poderão vir a ser implementados de forma semiautomática, usando os recursos lexicais já existentes para processamento computacional da língua.

Neste momento, procede-se à integração dos jogos elaborados no sistema VITHEA, complementando as funcionalidades já existentes, estando a explorar métodos mais ágeis de introdução dos exercícios nessa plataforma. Num futuro trabalho interdisciplinar pretendemos construir jogos com provérbios, gerados e corrigidos automaticamente, utilizando os recursos linguísticos já disponíveis.

Uma outra aplicação, de natureza didática, procurará explorar os contextos de utilização dos provérbios. Utilizar-se-ão técnicas de aprendizagem automática, nomeadamente árvores de decisão (Allen, 1995; Witten *et al.* 2011) para determinar, com base quer no léxico das frases envolventes, quer, se possível, no encadeamento retórico dessas frases (Pardo *et al.* 2004), os fatores que potenciam a utilização de um dado provérbio (ou suas variantes). Serão, então, selecionados os contextos em que esses fatores são mais relevantes para construir automaticamente um conjunto de exercícios a integrar numa plataforma de ensino de português assistido por computador (Pellegrini *et al.* 2012).

Referências bibliográficas:

Abad, A., Pompili, A., Costa, A., Trancoso, I., Fonseca, J., Leal, G., Farrajota, L. & Martins, I. (2013). Automatic word naming recognition for an on-line aphasia

- treatment system. *Computer Speech and Language*, Elsevier, 27 (6), 1235-1248.
- Basso, A. (1992). Prognostic factors in aphasia. *Aphasiology*, 6 (4), 337-348.
- Cazelato, S. (2003). *A interpretação de provérbios equivalentes por afásicos: um estudo enunciativo*. Dissertação de Mestrado em Linguística. IEL/UNICAMP.
- Correia, R.; Baptista, J, Nuno Mamede, Isabel Trancoso, Maxine Eskenazi, Automatic Generation of Cloze Question Distractors, In *Second Language Studies: Acquisition, Learning, Education and Technology, SLaTE: the ISCA SIG on Speech and Language Technology in Edu*, Waseda University, Tokyo, Japan, September 2010.
- Gangur, M. Automatic Generation of Cloze Questions. Proceedings of the 3rd International Conference on Computer Supported Education, Volume 1, Noordwijkerhout, Netherlands, 6-8 May, 2011
- Gorham, D. R. (1956). A proverbs test for clinical and experimental use. *Psychological Reports*, 2(Mono Suppl 1), 1-12.
- Kent, Raymond D. (ed). (2004). *The MIT Encyclopedia of Communication Disorders*. MIT Press.
- Léger, A., Démonet, J-F.; Ruff, S., Aithamon, B.; Touyeras, B., Puel, M., Boulanouar, K. & Cardebat, D. (2002). Neural substrates of spoken language rehabilitation in an aphasic patient: An fMRI study. *NeuroImage*, 17, 174-183.
- Marrafa, P. (2001) *WordNet do Português: uma base de dados de conhecimento linguístico*. Lisboa: Instituto Camões.
- Maziero, E.G., Pardo, T., Di Felippo, A. & Dias-da-Silva, B.C. (2008). A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. In *VI Workshop em Tecnologia da informação e da linguagem humana (TIL)* (Vila Velha, ES, Brasil, 28-29 Outubro 2008): 390-392.
- Mazzoni, M., Vista, M., Pardosi, L., Ávila, L., Bianchi, F. & Morew, P. (1995). Spontaneous evolution of aphasia after ischaemic stroke. *Aphasiology*, 6 (4), 387-396.
- Mendonça, V.; Coheur L. & Sardinha, A. (2015). VITHEA-Kids: a platform for improving language skills of children with Autism Spectrum Disorder. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'15)*, pp. 345-346.

- Morato, E. M. (2010). *A semiologia das afasias: perspectivas linguísticas*. São Paulo: Cortez.
- Musso, M., Weiller, C. Kiebel, S., Müller, S.P., Bülau, P. & Rijntjes, M. (1999). Training-induced brain plasticity in aphasia. *Brain*, 122, 1781-1790.
- Oliveira, H. G.; Santos, D.; Gomes, P. & Seco, N. (2008). PAPEL: a dictionary-based lexical ontology for Portuguese. In António Teixeira, *et al.* (eds.), *Computational Processing of the Portuguese Language*, (PROPOR 2008) Vol. 5190, (Aveiro, Portugal, 8-10 de Setembro de 2008), Springer Verlag: 31-40.
- Oliveira, H.G. & Gomes, P. (2014). ECO and Onto.PT: A flexible approach for creating a Portuguese wordnet automatically. In *Language Resources and Evaluation* 48(2): 373-39. Springer.
- Pianta, E., Bentivogli, L. & Girardi, C. (2002). MultiWordNet: developing an aligned multilingual database. In *Proceedings of the 1st International WordNet Conference*, January 21-25, 2002, Mysore, India: 293-302.
- Pies, R. D. (1994). *Clinical manual of psychiatric diagnosis and treatment: A biopsychosocial approach*. Washington DC.
- 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*, 5th-6th May, University of Algarve, Faro, Portugal, 427-446.
- Reis, S. & Baptista, J. (2016c). O uso de provérbios no ensino de português. In Soares, & Lauhakangas, Outi (eds.) *10th Interdisciplinary Colloquium on Proverbs, Actas ICP16 Proceedings*. Tavira: AIP-IAP, 2017. [no prelo]
- Reis, S. & Baptista, J. (2017a). Os provérbios em manuais de ensino de Português – Língua não materna. *Jornadas de Descrição do Português, in STIL-2017 - Symposium in Information and Human Language Technology* (JDP2017). Universidade Federal de Uberlândia, 2-4 outubro 2017 (aceite para publicação).

- Reis, S. & Baptista, J. (2017b). Estimating Lexical Availability of European Portuguese Proverbs. *Proceedings of EuroPhras2017*, Londres, 13-14 novembro 2017 (aceite para publicação).
- Santos, M.; Sougey, E. & Alchieri, J. (2009). Validity and reliability of the Screening Test for Alzheimer's Disease with Proverbs (STAD TAD P) for the elderly. *Arq Neuropsiquiatr*, 67 (3-B), 836-842.
- Sé, E. (2011). *Interpretação de provérbios por sujeitos com Doença de Alzheimer em fase inicial*. Tese de Doutorado. Universidade Estadual de Campinas.
- Simões, A. & Gómez Guinovart, X. (2014). Bootstrapping a Portuguese wordnet from Galician, Spanish and English wordnets. In *Advances in Speech and Language Technologies for Iberian Languages*, Proceedings of 2nd International Conference, IberSPEECH 2014, Las Palmas de Gran Canaria, Spain, volume 8854 of LNCS: 239-248. Springer.
- Small, S.L., Flores, D.K. & Noll, D.C. (1998). Different neural circuits subserve reading before and after therapy for acquired dyslexia. *Brain and Language*, 62, 298-308.
- Taylor, W. L. (1953). "Cloze procedure: A new tool for measuring readability". *Journalism Quarterly*. 30: 415-433.
- Peck, K.K., Moore, A.B., Crosson, B.A. Gaiefsky, M.; Gopinath, K.S.; White, K. & Briggs, R.W. (2004). Functional Magnetic Resonance imaging before and after aphasia therapy: shifts in hemodynamic time to peak during an overt language task. *Stroke*, 554-559.
- Bhogal, S.K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, 34 (4), 98-993.

A investigação para este artigo foi parcialmente financiada pela Fundação para a Ciência e a Tecnologia (ref. UID/CEC/50021/2013). A primeira autora agradece também o apoio do programa de Doutorado em Ciências da Linguagem da Universidade do Algarve.