

Os Provérbios em manuais de ensino de Português Língua Não Materna

Sónia Reis¹, Jorge Baptista^{1,2}

¹Faculdade de Ciências Humanas e Sociais – Universidade do Algarve (UALg)
Campus de Gambelas – 8005-138 – Faro – Portugal

²INESC-ID – Lisboa – Laboratório de Processamento da Língua Falada
Lisboa – Portugal

reis.soniamm@gmail.com, jrbaptis@ualg.pt

Abstract. *Proverbs present a wide variety of structures and may serve many communicative purposes. Due to their cultural and linguistic richness, they can be used to attain different didactic objectives, specifically in foreign language teaching. In this paper, we investigate how proverbs are in fact used in a corpus of Portuguese as Foreign Language (PFL) textbooks, using natural language processing (NLP) tools and resources. Results will be compared with previous findings on a corpus of Portuguese textbooks for native speakers.*

Resumo. *Os provérbios apresentam uma grande variedade de estruturas e podem servir diversos propósitos comunicativos. Devido à sua riqueza cultural e linguística, prestam-se ainda a múltiplos objetivos didáticos, nomeadamente no ensino de Português como Língua não Materna (PLNM). Neste trabalho, investigamos como são de facto utilizados os provérbios em manuais de PLNM, usando ferramentas e recursos de processamento computacional de linguagem natural (PLN). Os resultados são comparados com observações já feitas sobre um corpus de manuais de Português para falantes nativos.*

1. Introdução

Os provérbios apresentam uma grande diversidade de estruturas formais e servem uma multiplicidade de propósitos comunicativos (Charteris-Black, 1995; Hrisztova Gotthardt & Varga, 2015). A sua riqueza linguística e cultural faz deles um recurso bastante usado em ensino das línguas, uma vez que são frases concisas e linguisticamente muito diversificadas, de fácil acesso, podendo servir diferentes objetivos didáticos. Simultaneamente, são uma manifestação particularmente expressiva da cultura da comunidade linguística que os veicula e emprega numa grande variedade de situações comunicativas (Mieder, 2004; Arif & Abdullah, 2016; Salbego & Osborne, 2016). Relativamente ao ensino de português, os provérbios surgem quer nos manuais de Português Língua Materna (PLM), quer nos manuais de Português Língua não Materna (PLNM). Neste trabalho, pretendemos caracterizar como são empregues os provérbios num *corpus* de manuais atuais de PLNM, recorrendo a técnicas e recursos de processamento de linguagem natural (PLN).

2. Os Provérbios no Ensino de Línguas – Trabalhos Relacionados

O uso de provérbios pode contribuir para o ensino das línguas, tal como o preconizam, entre outros documentos ordenadores, no panorama europeu, o *Quadro Europeu Comum de Referência para as Línguas* (QERL) (Conselho da Europa, 2001) e o *Portfolio Europeu das Línguas* (PEL) (Schneider *et al.* 1999). Especificamente para a língua portuguesa, refira-se o documento orientador *Português Língua Não Materna no Currículo Nacional* (Leiria *et al.* 2008), onde são estabelecidos os princípios e as linhas orientadoras para a integração dos alunos dos diferentes níveis de ensino; bem como o *Referencial Camões – Português Língua Estrangeira* (Direção de Serviços de Língua e Cultura, 2017), um recurso disponibilizado de forma aberta ao diferente público que trabalha em ensino, aprendizagem, criação e avaliação de materiais de PLE. Especificamente sobre o uso de provérbios em ensino de língua, refira-se ainda algumas propostas de atividades dirigidas não só a alunos de PLM como também a alunos de PLNM (Martins, 2010; Pereira, 2015). Todavia, nestes trabalhos nem sempre são explícitos os critérios de seleção dos provérbios empregues ou são apresentadas as suas variantes mais usuais, aspetos que consideramos cruciais para uma boa aprendizagem, um conhecimento adequado e uso proficiente deste tipo de expressões.

3. Metodologia

Utilizámos um *corpus* constituído por 9 manuais de PLNM (v. lista apresentada depois das referências), com aproximadamente 250 mil palavras, abrangendo diferentes níveis de proficiência linguística (A1-B2)¹ do QERL e incluindo tanto manuais escolares como cadernos de exercícios, todos editados pela LIDEL². Os textos dos manuais foram digitalizados e convertidos em formato de texto simples para processamento automático.

A fim de determinar a distribuição de provérbios nos textos, desenvolveu-se um conjunto de recursos (biblioteca de grafos) (Reis & Baptista 2016a), baseados em técnicas de máquinas de estados finitos (Paumier, 2016), construídos a partir de uma base de dados com mais de 114 mil provérbios, digitalizados a partir de 4 coletâneas (ver lista depois das referências). Depois de removidas as palavras gramaticais (*stopwords*) consideradas não essenciais para a identificação dos provérbios, cada entrada da base de dados foi associada a uma chave, formada pelos elementos lexicais (*palavras-chave*) que constituem o provérbio, tipicamente nomes, verbos e adjetivos. Assim, por exemplo, para o provérbio *Olho por olho, dente por dente* [ID=PP_AM08501] obtemos a chave {olho-olho-dente-dente}. Ao todo, cerca de 52 mil chaves diferentes permitiram desde logo agrupar sob uma mesma *unidade paremiológica* (isto

¹ Não foi encontrada nenhuma coleção de manuais que apresentasse todos os níveis do QECR (A1-C2). Relativamente aos manuais *Português sem fronteiras* (PSF), usou-se as edições de 1997 (PSF1 e PSF2) e de 1999 (PSF3), por conveniência. As edições mais recentes (2009 e 2007, respetivamente para PSF1 e PSF2), já aparecem associadas a níveis do QECR, informação que usámos para este estudo. Ainda se encontra em preparação a nova edição do PSF 3.

² Esta opção deveu-se ao facto de esta editora ser uma das mais implementadas no sistema de ensino português.

3.2. Pesquisa por sequências de palavras-chave

A partir das chaves associadas aos provérbios da base de dados foram então construídos automaticamente um conjunto de transdutores, constituídos pelas palavras-chave de cada provérbio, admitindo um conjunto de inserções de 0 a 3 palavras e alguns sinais de pontuação entre cada palavra-chave. Foram construídos três tipos de transdutores diferentes: com 2, 3 e 4 palavras-chave. Não foram tidos em conta os 124 provérbios da base de dados com menos de 2 palavras-chave. Para os provérbios que contêm mais de 4 palavras-chave, usaram-se apenas as primeiras 4 palavras.

Este método não é tão preciso como o anterior. Por exemplo, não são consideradas as variações morfossintáticas das palavras-chave. Pela sua natureza, este método recupera muitas ocorrências espúrias, isto é, que não são efetivamente provérbios, mas permite capturar um grande número de expressões. Nomeadamente, permite capturar variações criativas destes provérbios como, por exemplo, certas inserções: *Quem te avisa (nem sempre!) teu amigo é*, embora não permita encontrar variantes que apresentem variação morfossintática, como sucede em *Quem te avisou, teu amigo foi*, ambos exemplos encontrados na internet. A Tabela 1 apresenta os resultados deste método.

Tabela 1. Resultados da pesquisa com os diferentes tipos de grafos

FST	Pal-chave	Tot. chaves	Chaves dif.	Concordâncias	Provérbios	UP. dif.
KW2	2	4.175	1.741	291	4	4
KW3	3	13.372	5.577	41	14	14
KW4	4	33.864	13.837	9	5	5
	+4	62.879	30.576	15	15	6
total	—	114.290	51.731	356	—	—

A coluna mais à esquerda indica o tipo de grafo. Os grafos são construídos tendo em conta o número de palavras-chave que formam a chave de cada provérbio. O grafo de tipo KW4 foi também usado para os provérbios com mais de 4 palavras-chave, utilizando apenas as 4 primeiras de cada chave. Segue-se o número total de chaves de cada tipo bem como o número de chaves diferentes usadas para construir os FST desse tipo. Após a verificação manual das concordâncias extraídas do *corpus*, identificaram-se os provérbios encontrados, para os quais se indica o número de unidades paremiológicas (UP) diferentes. Não obstante as limitações deste método, acima mencionadas, ele permitiu encontrar provérbios que não tinham sido captados pelo método anterior. Trata-se, concretamente, de um exercício em que se pedia para completar um provérbio truncado, e.g. *Mais (valer) prevenir do que remediar* e de um problema de digitalização num manual (*Casa onde não há pão, [...] todos ralham e ninguém tem razão*). Como algumas ocorrências dos provérbios encontrados emparelhavam com grafos diferentes. Assim, ao todo foram encontradas 27 ocorrências de provérbios, correspondentes a 14 unidades paremiológicas distintas.

3.3. Pesquisa por expressões introdutórias e lista de termos associados a provérbios

Repetindo a metodologia de Reis & Baptista (2016c), utilizada para a identificação automática de provérbios nos manuais de língua materna, procedeu-se ainda à pesquisa das expressões introdutórias associadas aos provérbios e com base dos termos a estes associados, que descrevemos agora. Este método consistiu na construção de

transdutores que descrevem expressões usadas para introduzir provérbios no discurso, tais como, *como diz o povo*, *lá dizia a minha avó*, *como se costuma dizer*, *nunca/já ouviram dizer que*. Ao todo, cerca de 15 expressões deste tipo foram representadas por transdutores de estados finitos. Acessoriamente, utilizou-se uma lista de 8 termos, usados por vezes como sinónimos de *provérbio*: *adágio*, *aforismo*, *anexim*, *axioma*, *ditado* (nome), *parémia*, *prolóquio*, *rifão*. Estes transdutores admitem variação morfossintática dos elementos das expressões introdutórias e dos termos pesquisados. Como resultado desta pesquisa obtivemos 18 ocorrências, das quais 15 continham provérbios, correspondendo a 14 UP diferentes, pois nem todas estão associadas a provérbios, uma vez que algumas destas são apenas menções destes termos. Na Figura 3, podemos observar os resultados desta concordância.

astronauta lápis rato erva faca [\[ditado\] \(term\)](#) homem inferno obra queijo urso bolo ovo jota s
ma enorme carga de água. 4. Bem [\[diz o povo \] \(introd\)](#): «Abril, águas mil». 5. O Steve matou s:
todo o dia seguinte choveu. Bem [\[diz o povo \] \(introd\)](#): «Abril, águas mil». No entanto nem tud:
o mesmo. Estou muito bem assim. [\[Nunca ouviram dizer que \] \(introd\)](#) «gordura é formosura»? Bom
asé incapaz de matar uma mosca. [\[Provérbio\] \(term\)](#): 6. Situação: Não gosto muito da cor do meu

Figura 3. Excerto da concordância da pesquisa de expressões introdutórias e lista de termos

Entre chavetas encontramos o termo ou expressão introdutória capturados e entre parênteses o tipo respetivo (*term*=termo, *introd*=expressão introdutória). Este método apresenta a desvantagem resultante de nem sempre os termos ou expressões introdutoras estarem efetivamente associados a provérbios. A pesquisa permitiu encontrar 8 provérbios diretamente associados a estes termos/expressões e 7 provérbios no contexto circundante. Destas 15 ocorrências, 7 correspondem a provérbios (6 unidades paremiológicas diferentes) que não tinham sido capturados pelos métodos anteriores.

4. Resultados

Os diferentes métodos de pesquisa permitiram encontrar 34 provérbios no *corpus* de manuais escolares selecionados, correspondendo a 20 unidades paremiológicas distintas. Na Tabela 2 apresentamos o número de provérbios encontrado por manual, bem como o género textual em que ocorrem.

Tabela 2. Distribuição dos provérbios pelos manuais escolares

Manuais	Nível QCERL	Exercícios	Títulos	Corpo texto	Total prov
NOP 1	A1/A2	0	1	0	1
NOP 2	B1	1	0	1	2
NOP 3	B2	7	2	5	14
NOP1 - CE	A1/A2	0	1	2	3
NOP2 - CE	B1	0	0	0	0
NOP3 - CE	B2	0	0	1	1
PSF 1	[A1/A2]	1	0	1	2
PSF 2	[B1]	0	0	0	0
PSF 3	[B2]	7	0	4	11
Total		16	4	14	34

Legenda: NOP – *Na Onda do Português*; CE – *Caderno de Exercícios*;
PSF – *Português sem Fronteiras* (v. referências no final do artigo).

Praticamente em todos os manuais se encontraram provérbios, excepto em dois, ambos correspondendo ao nível intermédio B1. Note-se que a ocorrência de provérbios em manuais deste nível também é escassa (2). Como se pode verificar, o número de

provérbios encontrados nos manuais pertencentes ao nível B2 (26) é muito superior ao número de provérbios nos manuais de níveis A1/A2 (6) e corresponde a $\frac{3}{4}$ do número total de provérbios identificados. Este era um resultado esperado, uma vez que o *Quadro Europeu Comum de Referência para as Línguas (QERL)* determina que:

A partir do nível B2, considera-se que os utilizadores são capazes de se exprimir adequadamente numa linguagem que é sociolinguisticamente apropriada às situações e aos interlocutores e que começam a adquirir a capacidade de enfrentar a variação do discurso, além de possuírem, em elevado grau, domínio do registo e das *expressões*.

(Conselho da Europa, 2001: 172, itálico nosso)

Realce-se que é apenas nos níveis C1 e C2 que o mesmo *Quadro* prevê que os utilizadores “sejam capazes de reconhecer um vasto leque de expressões idiomáticas e de coloquialismos” (*idem*: 173, nível C1) e de possuir “um bom domínio de expressões idiomáticas e de coloquialismos com consciência dos níveis conotativos do significado” (*idem*: *ibidem*, nível C2).

Para além da distribuição dos provérbios nestes manuais, quisemos determinar o género textual em que ocorriam, distinguindo exercícios, corpo do texto (para leitura) e títulos. Verificámos que 16 (47%) dos provérbios encontrados se encontram em exercícios, 14 (41%) no corpo do texto e 4 (12%) em títulos. O tipo de exercícios encontrado consiste essencialmente em: (i) escolher o provérbio que melhor se adequa a uma determinada situação, (ii) explicar o sentido do provérbio, (iii) justificar se está de acordo com a afirmação do provérbio, (iv) completar as palavras em falta no provérbio e (v) dizer se conhece algum outro provérbio para além do apresentado. Verificámos que num dos exercícios, cujo objetivo era completar as palavras em falta do provérbio, qualquer frase serviria o mesmo propósito, apesar de a escolha dos autores ter recaído sobre este tipo de expressões. Observámos ainda que apenas 7 provérbios se repetiram nos manuais analisados e 13 provérbios surgiram 1 só vez. Desta forma, não foi possível determinar se alguns provérbios seriam mais usuais nestes manuais, até porque alguns dos provérbios que se repetiram surgem no mesmo manual, embora por vezes em diferentes situações de uso.

Em Reis & Baptista (2017), procurou-se determinar a disponibilidade lexical dos provérbios, isto é, quais são os mais usuais em português europeu. Para tal, usou-se uma amostragem estratificada dos provérbios da base de dados e posterior aplicação de questionário a falantes nativos (mais de 700 respostas). Estes dados foram também confrontados com a frequência desses mesmos provérbios na internet, usando dois motores de busca distintos e ainda comparados com a sua frequência num *corpus* de texto jornalístico. Constituiu-se, assim, uma lista de 276 provérbios bastante usuais (nível 2) e outros 566 menos usuais mas ainda bem conhecidos da generalidade dos falantes (nível 1). Esta lista servirá de referência para diversas observações sobre o uso de provérbios em textos. Assim, usando como referência estes dados de disponibilidade lexical, verifica-se que, das 20 unidades paremiológicas encontradas, 14 são considerados bastante usuais (nível 2) e 6 são apenas usuais (nível 1). Relativamente aos provérbios que surgiram mais vezes, observa-se a seguinte distribuição por níveis de disponibilidade: *Querer é poder* (nível 2) e *Recordar é viver* (nível 1) com 4 ocorrências cada um; *O fruto proibido é sempre o mais apetecido* (nível 2), *Mais vale prevenir do que remediar* (nível 2) e *Casa onde não há pão, todos ralham e ninguém tem razão* (nível 1), com 3 ocorrências, e ainda *Abril, águas mil* (nível 2) e *Da discussão nasce a*

luz (nível 1), com 2 ocorrências. Apenas um provérbio (*Quem conta um conto um conto acrescenta um ponto*) aparece tanto nos manuais de PLNM (1 ocorrência) como de PLM (7 vezes, sendo até o mais frequente). Verificámos também que os tipos de exercícios usados nos manuais de PLNM são os mesmos que se encontram nos manuais de PLM.

5. Conclusões

Após a análise das expressões recuperadas por cada um dos três métodos acima descritos, é possível afirmar que estes se complementam entre si, permitindo a identificação de expressões proverbiais em textos. Naturalmente, o método dos transdutores das unidades paremiológicas é muito mais preciso, mas requer a prévia formalização das variantes e sua reunião num mesmo grafo, o que é um trabalho moroso e demorado. Poderá ser, no entanto, um método adequado para os 842 provérbios de uso frequente, tal como identificados por Reis e Baptista (2017), reservando os outros dois métodos, de carácter mais exploratório, para os restantes casos. É, no entanto, possível ainda aperfeiçoar o método das palavras-chave, tornando-o mais preciso, nomeadamente, recorrendo à pesquisa dessas palavras pelo respetivo lema, limitando as inserções no caso de palavras-chave adjacentes.

Com base no *corpus* em análise, este estudo permitiu verificar que os provérbios são efetivamente usados nos manuais de PLNM e estão maioritariamente associados aos níveis de proficiência mais elevados (B2). Note-se que nenhum dos manuais em análise pertencia aos níveis C1 e C2, nível para o qual se preconiza, nos documentos ordenadores de ensino de PLE, a aquisição de competências relativamente a este tipo de expressões. À semelhança dos manuais de PLM, os provérbios surgem com mais frequência em exercícios do que no corpo dos textos. Os tipos de exercícios encontrados nos manuais de PLNM são idênticos aos dos manuais de PLM, mas nestes surgem sobretudo nos 5º e 6º anos (2º ciclo) do Ensino Básico. Alguns dos exercícios que apresentam provérbios poderiam perfeitamente apresentar qualquer outro tipo de frase considerando o objetivo pedagógico a atingir, pelo que a escolha de expressões proverbiais para estes exercícios parece pouco fundamentada. Todos os provérbios que constam destes manuais podem ser considerados usuais, com base em dados de disponibilidade lexical (Reis e Baptista 2017).

Referências

- Arif, M. and Abdullah, I. (2016). The impact of output communication on EFL learners' metaphor second language acquisition. *Social Sciences* (Pakistan) 11:9, p. 1940-1947.
- Charteris-Black, J. (1995). Proverbs in communication. *Journal of Multilingual and Multicultural Development*, 16:4, p. 259-268.
- Conselho da Europa. (2001). *Quadro Europeu Comum de Referência para as Línguas: Aprendizagem, ensino, avaliação*. Edições ASA, Porto.
- Direção de Serviços de Língua e Cultura (2017) *Referencial Camões PLE – Português Língua Estrangeira*, Camões, Instituto da Cooperação e da Língua I.P., Lisboa.
- Hrisztova-Gotthardt, H. and Varga, M. (eds.). (2015). *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*. De Gruyter, Berlin.

- Leiria, I. (coord.); Martins, A.; Cordas, J.; Mouta, M. and Henriques, R. (2008). *Orientações Programáticas de Português Língua Não Materna (PLNM) Ensino Secundário*. Ministério da Educação.
- Martins, P. (2010). Do provérbio em contexto didático: proposta de trabalho. *Paremia*, 19: 2010, p. 93-102.
- Mieder, W. (2004). *Proverbs: A Handbook*. Greenwood Press, London.
- Paumier, S. (2016). *Unitex 3.1 – User Manual*. Université de Paris-Est/Marne-la-Vallée - Institut Gaspard Monge, Noisy-Champs.
- Pereira, A. (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. and 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, p. 208-217.
- Reis, S. and 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, p. 427-446.
- Reis, S. and 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. and Baptista, J. (2017). O provérbio como estímulo num terapeuta virtual. *VI Simpósio Mundial de Estudos sobre o Português (SIMELP)*, Simpósio 77, A Importância da Aprendizagem Lexical, Santarém, Escola Superior de Educação, Instituto Politécnico de Santarém (2017, aceite para publicação).
- Salbego, N. and Osborne, D. (2016). Schema activation through pre-reading activities: teaching proverbs in L2. *BELT – Brazilian English Language Teaching Journal* 7.2, p. 175-188.
- Schneider, G., North, B., Flügel, Ch. and Koch, L. (1999). *European Language Portfolio*. Berna: EDK. (<http://www.unifr.ch/ids/portfolio>)

Coletâneas de provérbios incluídas no corpus

- Costa, J. (1999). *O Livro dos Provérbios Portugueses*. Lisboa, Ed. Presença.
- Machado, J.P. (1996). *O Grande Livro dos Provérbios*. 1^a edição, Lisboa, Ed. Notícias.
- Moreira, A. (1996). *Provérbios Portugueses*. Lisboa, Ed. Notícias.
- Parente, S. (2005). *O Livro dos Provérbios*. 1^a edição. Lisboa, Ed. Âncora.

Lista de manuais do corpus de PLNM

- Coimbra, I. and Coimbra O. (1997). *Português sem Fronteiras 1* (nova versão revista e actualizada) e *Português sem Fronteiras 2*. Lidel, Lisboa.
- Leite, I. and Coimbra O. (1999). *Português sem Fronteiras 3*. Lidel, Lisboa.
- Ferreira, A. and Bayan H. (2011). *Na Onda do Português 1 e Na Onda do Português 2* (Manual e Caderno de Exercícios), 2^a edição. Lidel, Lisboa.

Ferreira, A. and Bayan H. (2012). *Na Onda do Português 3* (Manual e Caderno de Exercícios). Lidel, Lisboa.

Agradecimentos

Parte desta investigação foi suportada por fundos públicos, através da Fundação para a Ciência e a Tecnologia, Portugal (ref. UID/CEC/50021/2013).